Rhox4: Duplication, Evolution and Thymus Organogenesis

Lucy Morris

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

January 2006

Abstract

The *Rhox* family of homeobox-containing genes maps to three gene dense clusters, α, β and γ, on the mouse X chromosome. *Rhox* genes are primarily expressed in reproductive and extra-embryonic tissues and are reportedly regulated by co-linear mechanisms. *Rhox4* is also expressed in the developing and adult thymus, where its striking expression pattern has suggested a role in lineage specification/early organogenesis. The aim of this study was to investigate the role of *Rhox4* in thymus development. This was addressed through analysis of *Rhox4* expression in normal mouse development, and in mutant mice with known defects in thymus development; and detailed physical and expression mapping of the *Rhox4* locus. In addition, initial development of an RNAi strategy for evaluation of *Rhox4* function is described.

The main findings of this thesis were that *Rhox4* expression is restricted to the ventral region of the third pharyngeal pouch consistent with a role in thymic lineage specification. However, no changes in *Rhox4* expression were detected in mice with defects in thymus organogenesis, placing it upstream or outside of established transcriptional pathways. Physical mapping of the *Rhox4* locus identified unreported duplication events within the *Rhox* α locus that contains seven copies of *Rhox4* and eight copies of *Rhox2* and *Rhox3* in tandem array. I show that all seven *Rhox4* duplicates are expressed, although preferential expression occurs and differs between tissues. In contrast to reproductive tissues, I found no evidence of co-linear expression of the *Rhox* α cluster during thymus development. All sub-species of mice examined contained multiple copies of *Rhox4*. However, the only predicted orthologue of *Rhox4*, rat *Rhox4*, is present in a single copy, suggesting the duplications arose at the time of rat and mouse lineage divergence. Although no expression of rat *Rhox4* was detected in the embryonic or adult thymus, *Rhox3* expression was detected throughout thymus ontogeny, suggesting rat *Rhox3* as the functional orthologue of mouse *Rhox4*. Surprisingly, neither human *Rhox* gene was expressed in the embryonic thymus, indicating that either thymic *Rhox* expression

arose in a recent murine ancestor or human family members have lost this expression domain. No other differences in embryonic thymic transcription factor expression were observed between mouse and human, suggesting conservation of pathways. Together, these analyses suggest that *Rhox4* has a murine-specific role in thymus organogenesis. Finally, I describe an RNAi strategy for determining the role of *Rhox4* in the thymus.

Acknowledgments

Firstly, I would like to thank Clare for giving me the opportunity to work on *Rhox4,* I have learnt a huge amount and it has certainly kept me amused! I would also like to thank her for all the recent work put into reading my thesis. Thanks to Andrew for his help and advice on my original, but now obsolete, construct building efforts, and on gene conversion. Everybody in Clare's lab have been great, particularly thanks to Julie for the tea and talk, Terri for putting up with me and lots of help, Al for ceilidhs and embryos, Natalie and Marianne for very useful bits and pieces, Craig for help with lots of things, and Julie G for teaching me *in situs.* Many other people in the University have given me help and advice – I would like to mention Prof. Deborah Charlesworth for extremely helpful discussions on evolution, Tilo for RNAi advice and plasmids, Val for endoderm dissections and Ron for sectioning my *in situs*.

During my PhD I attended two courses, the Cold Spring Harbor Labs - *Molecular Embryology of the Mouse,* and the Wellcome Trust - *Working with the Human Genome Sequence.* Both courses were extremely informative and provided me with many of the skills and understanding that has gone into the work presented in this thesis. They were organised and attended by really fun, and interesting people, all of whom I would like to thank. My attendance on these courses, as well as two conferences would not have been possible without the financial support of the British Society for Developmental Biology, Boehringer-Ingelheim Fonds and the Cold Spring Harbor Labs. Additionally, and very importantly, thanks to the Wellcome Trust for supporting my PhD work.

Finally I would like to thank my family and friends: My parents for endless sympathy and vegetables, Daniel, Ol, Anne-Marie, Helen, Ben, Anna, Nicola, Shona, and of course Ian for his love and inspiration, and rationalisation of all difficulties encountered!

Content

Abbreviations

BAC – bacterial artificial chromosome

BLAST – basic local alignment search tool

Bmp – bone morphogenetic protein

bp – base pair

BSA – bovine serum albumin

cDNA – complementary deoxyribonucleic acid

DIG – deoxygenin

DN – double negative

DNA – deoxyribonucleic acid

DP – double positive

dsRNA – double-stranded RNA

E – embryonic day

EB – embryoid body

EMBL – the European molecular biology laboratory

ES – embryonic stem cell

Fgf – fibroblast growth factor

Gcm – glial cells missing

GFP – green fluorescent protein

Hox – homeobox

IL – interleukin

kb – kilabase basepair

LIF- leukaemia inhibitory factor

MHC – major histocompatibility complex

mRNA – messenger ribonucleic acid

miRNA – micro-RNA

MTS – mouse thymic stromal

NCBI – National Centre for Biotechnology Information

ORF – open reading frame

PCR – polymerase chain reaction

PTH – parathyroid hormone

RA – retinoic acid

RALDH – retinaldehyde dehydrogenase

RAR – retinoic acid receptor

RACE – rapid amplification of cDNA ends

Rhox – reproductive homeobox X-linked

RISC – RNA-induced silencing complex

RNA – ribonucleic acid

RNAi – RNA interference

RT-PCR – reverse transcriptase polymerase chain reaction

RXR – retinoid X receptor

shRNA – short hairpin RNAs

siRNA – small-interfering RNAs

Shh – sonic hedgehog

TCR – T cell receptor

TEC – thymic epithelial cell

UTR – untranslated region

Chapter 1. Introduction

1.1. Introduction to the thymus

The vertebrate thymus is a complex network of epithelial and lymphoid cells that are responsible for the generation of a self-restricted and self-tolerant T cell repertoire (Anderson and Jenkinson 2001). An intricate three-dimensional meshwork of thymic epithelial cells (TEC) forms the structural framework of the thymus (van Ewijk, Wang et al. 1999). T cell maturation occurs concomitant with migration through this complex epithelial structure, and TEC play a critical role in the T cell maturation process. The absence of TEC results in severe immunodeficiency due to the central role of T cells in the adaptive immune response. This is seen in humans (Frank, Pignata et al. 1999) and mice (Flanagan 1966) homozygous for naturally occurring mutations in *Foxn1*, a transcription factor that is required cell-autonomously for TEC maturation (Blackburn, Augustine et al. 1996; Nehls, Kyewski et al. 1996). The thymus is divided by histology into cortical and medullary regions. Medullary and cortical TEC are phenotypically and morphologically distinct, and play different roles in thymocyte development (Anderson and Jenkinson 2001; Gotter and Kyewski 2004).

1.2. Cellular and morphological events involved in thymus organogenesis

Thymus organogenesis involves formation of a common thymus/parathyroid primordium; specification of the thymic domain of the 3rd pharyngeal pouch; outgrowth, followed by colonisation of the thymic rudiment by T cell progenitors; and proliferation and maturation of thymic epithelial cells (Figure 1.1) (Blackburn and Manley 2004). Separation of the common primordia into separate thymus and parathyroid rudiments occurs between E13 – E13.5, following which the two thymic primordia migrate caudo-ventrally and fuse at the midline to form the mature organ (Figure 1.2) (Schreier and Hamilton 1952; Hammond 1954).

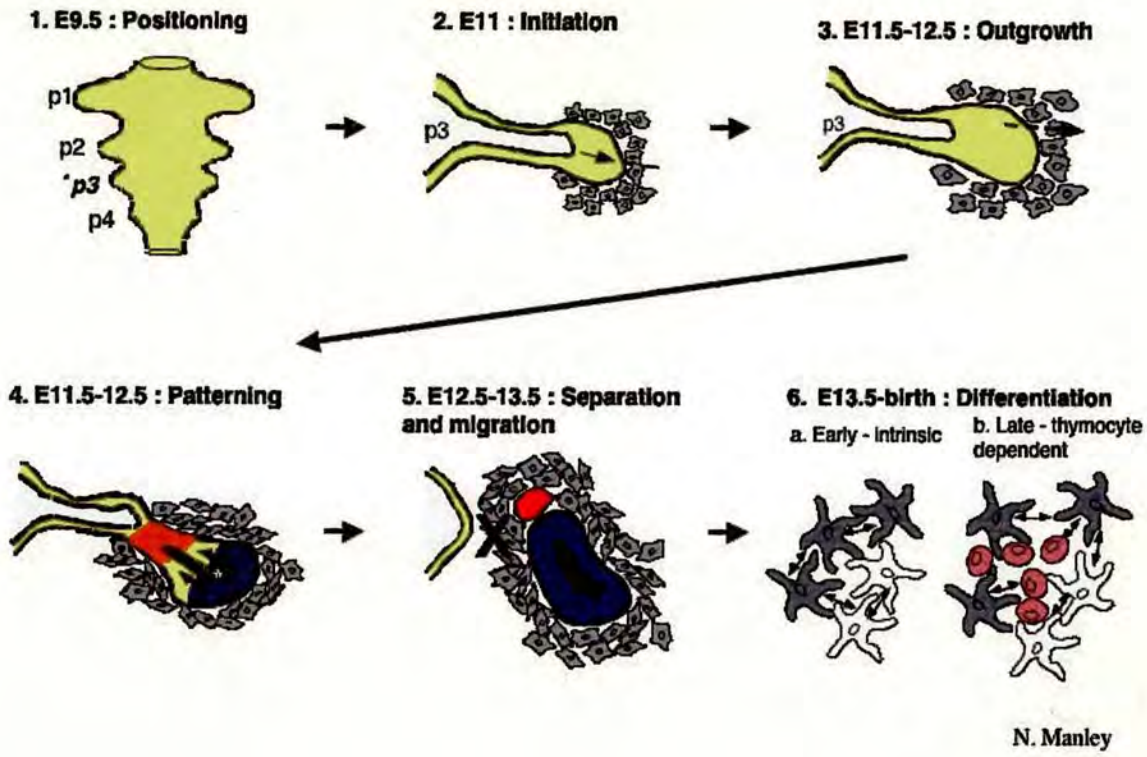


Figure 1.1 Summary of morphological events involved in thymus organogenesis. 1) The 3rd pharyngeal pouch forms around E9.5; 2) Condensation of mesenchymal cells around the pouch occurs from E11 and 3) the pouch begins to grow away from the main gut tube, T cell precursors also begin to arrive at this stage; 4) The thymic (blue) and parathyroid (red) regions of the 3rd pharyngeal pouch can be identified by expression of *Foxn1* (thymus) and *Gcm2* (parathyroid); 5) Separation of the common primordia occurs by E13.5 and the thymus then undergoes proliferation and differentiation to form the mature organ (6). Diagram adapted from Blackburn and Manley (2004).

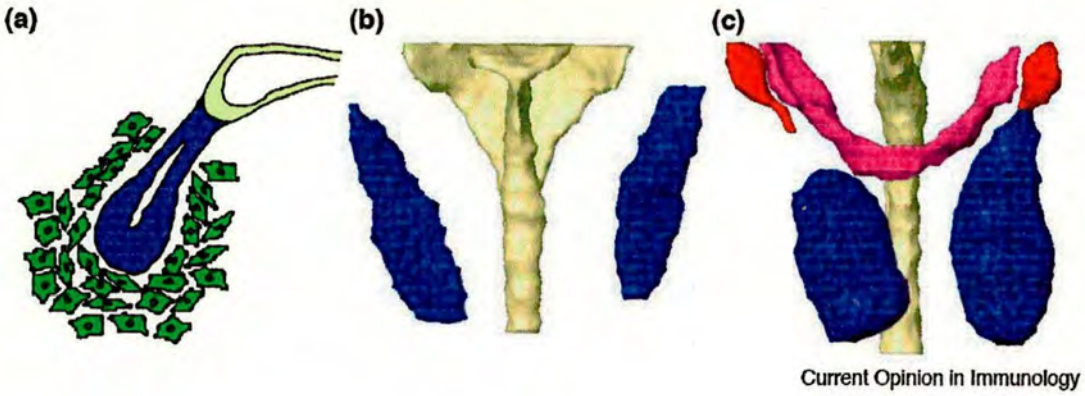


Figure 1.2 Stages of early thymus and parathyroid development. a) Diagram of an E11.5 parasaggital section, showing the common thymus/parathyroid primordium (blue) budding off the left 3rd pharyngeal pouch. The primordium is surrounded by mesenchyme (green). b, c) Frontal views of three-dimensional reconstructions showing the positions of the thymic/parathyroid primordia. At E12.5 the common primordia have separated from the pharynx (beige) (b). By E13.5 (c), the parathyroid (red) and thymic (blue) domains separate. The position of the thyroid gland is shown in pink. Taken from Manley and Blackburn 2003.

1.2.1. Components of the pharyngeal region

Gastulation results in the formation of three germ layers, ectoderm, endoderm and mesoderm (Hogan 1999). The ectoderm and endoderm are simple epithelial layers bounding the mesodermal germ layer. The pharyngeal region consists of mesodermal pharyngeal arches, which are evident as bilateral bulges on the ventral side of the embryo, pockets of foregut endoderm called the pharyngeal pouches, and surface ectodermal invaginations, termed clefts (Figure 1.3) (Graham and Smith 2001). There are five pharyngeal arches in mouse, which are separated by opposing pharyngeal pouches and clefts, giving rise to a structure that is segmented down the anterior-posterior axis. The pharyngeal region gives rise to many of the components of the jaw and neck, including the thymus.

1.2.1.1 Pharyngeal arches

The pharyngeal arches contain both neural crest and non-neural crest derived mesenchyme. These cell types directly give rise to the facial and neck skeletal elements and muscles and additionally contribute to organs derived from the pharyngeal endoderm, including the developing thymus. Neural crest is a transient, migratory cell population formed between the neural tube and the surface ectoderm. Neural crest derived mesenchyme migrates into the pharyngeal region from E9.0, and was initially believed to carry patterning information to the pharyngeal region from the segmented hindbrain (Noden 1983). However, ablation of neural crest cells in chick embryos does not affect patterning of the pharyngeal endoderm (Veitch, Begbie et al. 1999) and endoderm grafting experiments demonstrate that the Hox-negative neural crest, which populates the first pharyngeal arch, is plastic and responds to endodermally derived signals (Couly, Grapin-Botton et al. 1998; Couly, Creuzet et al. 2002). Additionally, evolutionary evidence indicates that the pharyngeal system arose before neural crest, suggesting that patterning of the pharyngeal region is likely to be neural crest independent (Veitch, Begbie et al. 1999; Graham and Smith 2001). However, neural crest cells do appear to play a central role in species-specific skeletal patterning in ducks

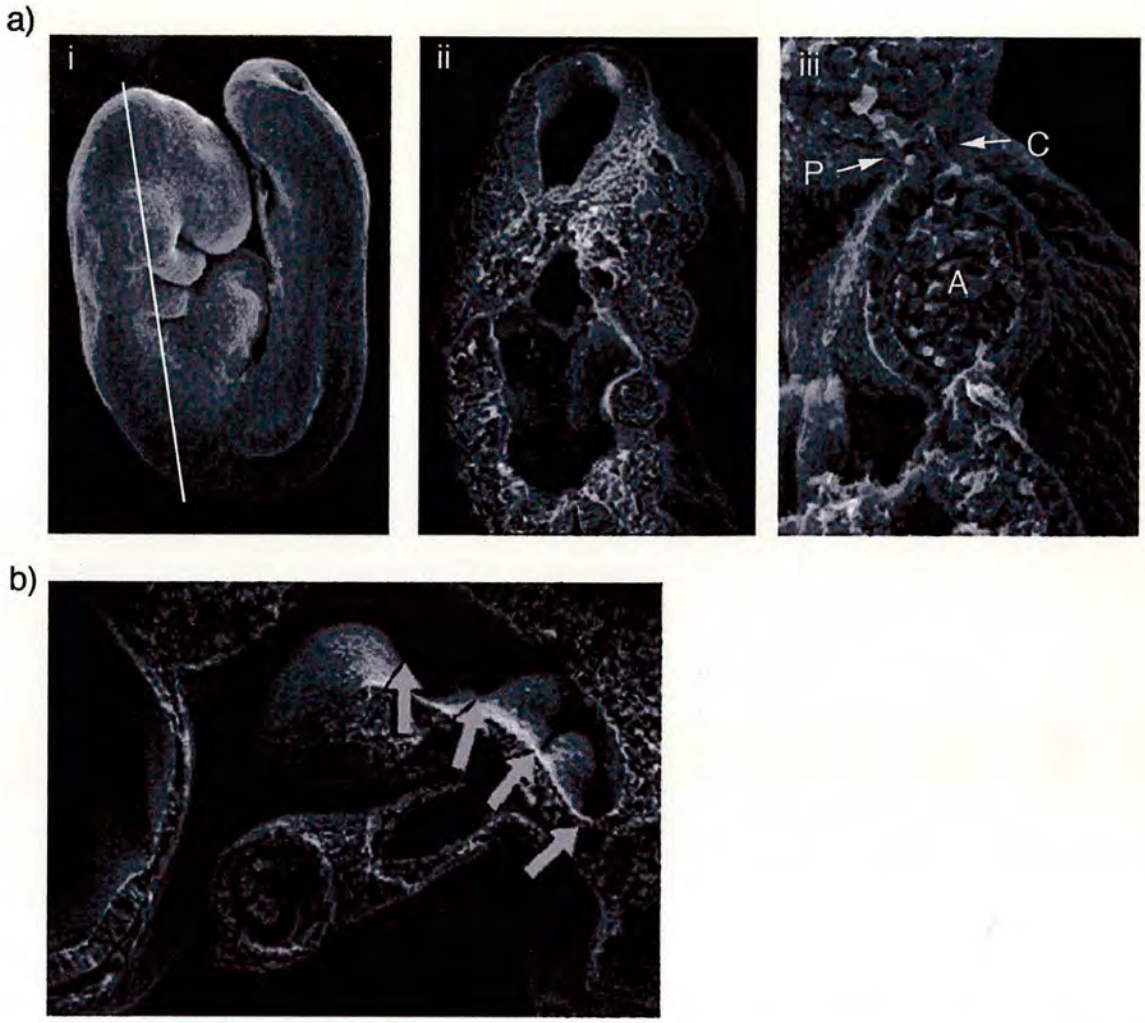


Figure 1.3 Structure of the pharyngeal region. Scanning electron microscopy of; ai) whole E9.0 embryo, aii) coronally bisected E9.0 embryo, revealing the structures present in the pharyngeal region, and aiii) higher power image of coronally bisected E9.0 embryo, showing details of apposition of the pharyngeal pouches and the pharyngeal clefts. P – endodermal pharyngeal pouch, A – mesenchymal pharyngeal arch and C – ectodermal pharyngeal cleft. Orientation, head up; b) E9.5 embryo hemisected to reveal the four pharyngeal pouches (arrows). Orientation, head is left, dorsal is up. Images taken from www.med.unc.edu/embryo_images/.

and quails, as demonstrated by inter-species grafting experiments (Tucker and Lumsden 2004).

1.2.1.2 The pharyngeal pouches

The pharyngeal pouches are initially evident as bilateral buds, present at distinct positions down the anterior-posterior axis of the foregut endoderm (Figure 1.3). Their formation segments the pharyngeal region and is necessary for pharyngeal arch formation, probably due both to architectural constraints and the production of signalling molecules. It is also evident that the pharyngeal endoderm plays a primary role in patterning of the pharyngeal region, as no arch formation occurs in the absence of endoderm (Alexander, Rothenberg et al. 1999) and pouch defects result in the malformation of the corresponding arch and ectoderm derivatives (Crump, Maves et al. 2004; Crump, Swartz et al. 2004; Haworth, Healy et al. 2004). Structures arising from the pharyngeal pouches include components of the tonsils, eustachian tubes, thyroid, thymus and parathyroid glands, all of which eventually bud off from the main gut tube and migrate to distinct locations within the embryo. Opposing each pouch there is an ectodermal invagination, or pharyngeal cleft, which is also necessary for shaping the pharyngeal region and an important source of signals (Macatee, Hammond et al. 2003).

The mechanism via which the pharyngeal pouches form is poorly understood. However, live imaging of zebrafish embryos suggests that cells in distinct domains of the endoderm actively migrate away from the gut tube. The migration maybe mediated by a chemotactic signal, as cytoplasmic processes can be seen extending laterally (Crump, Maves et al. 2004). The characteristic slit shape of the pouches is maintained by actin cables present on the apical side of pouch cells, which are linked via adherins junctions (Quinlan, Martin et al. 2004). When actin polymerisation is inhibited in whole embryo culture, pouch development continues but, due to the lack of rigidity, the sheet of pharyngeal pouch endoderm forms bulges and folds (Quinlan, Martin et al. 2004).

1.2.2. Morphological and cellular events in thymus organogenesis

The thymus arises as a result of interactions between 3rd pharyngeal pouch endoderm and surrounding neural crest derived mesenchyme. Although the pharyngeal cleft ectoderm has also been reported to also have a role in formation of the thymic primordium (Cordier and Haumont 1980), grafting experiments have shown that the endoderm alone is sufficient in both chick (Le Douarin and Jotereau 1975) and mouse (Gordon, Wilson et al. 2004). The 3rd pharyngeal pouch is formed around E9.5 and gives rise to a common thymus and parathyroid primordium (Schreier and Hamilton 1952; Hammond 1954).

The ventral region of the 3rd pharyngeal pouch gives rise to the thymus and the dorsal portion to the parathyroid gland. The complimentary thymic and parathyroid regions of the 3rd pouch can be identified from E11.25 by *in situ* hybridisation for the domain-specific transcription factors, *Foxn1* in the thymic domain and *Gcm2* in the parathyroid domain (Gordon, Bennett et al. 2001). However, specification of the 3rd pharyngeal pouch into thymus and parathyroid regions may occur at, or before the time of pouch formation as grafting of the 2nd and 3rd pouch region of E8.5 – E9.0 foregut endoderm under the kidney capsule of nude mice results in the formation of a functional thymus, although no parathyroid structures were reported in this study (Gordon, Wilson et al. 2004). Additionally, *Gcm2* is expressed in the dorsal region of the pouch from E9.5, at the time of pouch formation (Gordon et al. 2001).

Two monoclonal antibodies, MTS20 and MTS24 mark the E9.5 pharyngeal endoderm, and subsequently identify a population of thymic epithelial cells, which is common in the early embryonic thymus but very rare in adult (Bennett, Farley et al. 2002; Gill, Malin et al. 2002). Grafting of reaggregate cultures of E12.5 or E15.5 $MTS20^{+}$ 24^{+} thymic epithelial cells and mesenchymal cells under the kidney capsule of nude mice, gave rise to a functional thymus (Bennett, Farley et al. 2002; Gill, Malin et al. 2002). This work suggests that embryonic thymic epithelial cells

differentiate from a common progenitor cell, which is defined by staining for MTS20 and 24.

A mesenchymal-derived capsule condenses onto the common thymus/parathyroid primordium by E11.5, and this surrounding neural crest-derived mesenchyme invades the thymic rudiment at around E13. Expression and transgenic studies have indicated that signalling between the mesenchyme and the thymic rudiments occurs, and may be required for epithelial cell differentiation and the early steps in thymopoiesis (Anderson, Anderson et al. 1997; Shinohara and Honjo 1997; Jiang, Rowitch et al. 2000; Suniara, Jenkinson et al. 2000; Revest, Suniara et al. 2001; Abu-Issa, Smyth et al. 2002). However, although genetic marking of neural crest-derived cells suggests that they do not contribute significantly to the mature organ (Jiang, Rowitch et al. 2000; Yamazaki, Sakata et al. 2005) these analyses are likely due to transgene silencing (M. Coles, NIMR, London, UK, personal communication). The thymus is closely associated with the vascular system and vascularisation may also play a role in organisation of the medullary region (Anderson, Anderson et al. 2000).

Early foetal lymphocyte precursors migrate to the thymic primordia through the connective tissue, and are first detected in the thymus around E11 (Moore and Owen 1967; Le Douarin and Jotereau 1975). In adult, lymphocyte precursors migrate from the bone marrow into the circulation, enter the thymus at the cortico-medullary junction and then move to the sub-capsular region of the cortex (Lind, Prockop et al. 2001). Whilst migrating through the cortical region, thymocytes undergo selection - termed positive selection - for propensity to bind self-MHC (Jameson, Hogquist et al. 1995); whereby cells that fail to engage self-MHC die by apoptosis due to a lack of stimulatory signals. This removes cells bearing non-productively rearranged T-cell receptor genes (TCR), importantly including those in which receptor rearrangement has generated translocations. Positively selected cells then pass into the medullary regions of the thymus, where they undergo negative selection (Nossal 1994). Any thymocyte that binds antigen with high affinity/avidity, either on thymic epithelial cells (TEC) or dendritic cells, is eliminated.

Defined cortical and medullary regions of the thymic rudiment can be identified from E12.5 by differentiation marker expression (Klug, Carter et al. 1998; Bennett, Farley et al. 2002; Klug, Carter et al. 2002). However, expansion and maintenance of cortical and medullary compartments requires interactions with lymphocyte precursors. Mice bearing mutations that cause blocks in T cell development also result in aberrant thymic development. These include, hCD3ε line 26 transgenic mice in which the insertion of the transgene causes lymphocyte precursors to arrest before commitment to the T cell lineage has occurred (Hollander, Wang et al. 1995). In these mice the thymus is cystic, contains a disorganised cortex and no medullary epithelial cells are present (Hollander, Wang et al. 1995; Wang, Hollander et al. 1996; Klug, Carter et al. 1998). Arrest of lymphocyte development immediately proceeding commitment to the T cell lineage, as seen in $RAG1^{-/-}$ mice, results in the production of an organised cortex but no medulla (Hollander, Wang et al. 1995; Klug, Carter et al. 1998). Just as T cells are necessary for TEC development, fully functional TEC are required to fully support T cell development (See 1.3) (Wallin, Eibel et al. 1996; Su, Ellis et al. 2001; Hetzer-Egger, Schorpp et al. 2002; Su and Manley 2002).

1.3. Genetic control of thymus organogenesis

Many mouse mutants have now been identified that have defects in thymus organogenesis (Table 1.1). These genetic studies are beginning to unravel transcription and signalling molecule pathways involved in formation of the thymic primordia and thymus organogenesis. Not surprisingly, many of these factors are also involved in other developmental systems, including, *Hox, Pax, Eya* transcription family members, and Shh, Fgf, Wnt and Bmp signalling molecules. Interpretation of many of the thymic mutant phenotypes is not simple, due to the complex cellular interactions that take place during thymus development. Additionally, the study of many early thymic mutants is hampered by embryonic or neonatal lethal phenotypes, as the pharyngeal region, including the endoderm, plays a central role in development of the respiratory and circulatory systems.

Table 1.1 Thymic phenotypes of mutant mice with defects in thymus organogenesis. PP – pharyngeal pouch, PA – pharyngeal arch, PC – pharyngeal ectodermal clefts, PA – pharyngeal arch.

Gene knockout/hypomorphic allele	Gene expression	Phenotype	Expression of other genes in transgenic	Possible role of gene in thymus organogenesis
***Chordin*$^{-/-}$**	Node and notochord at E7.0 and subsequently the dorsal endoderm and restricted to the pharynx E8.5	2nd to 4th PP and PA fail to form	*Tbx1, Fgf8* and *Pax1* expression down-regulated	Required for PP and PA formation possibly via regulation of *Tbx1*
***Tbx1*$^{-/-}$**	Foregut endoderm from E8.0 then all 4 PPs, non-NC mesenchyme and surface ectoderm	2nd to 4th PP severe hypoplasia, athymia (except 1 embryo) Abnormal NC migration	Down-regulation of *Gcm2, Pax1, Pax9, Fgf8*	Required for PP formation and segmentation Later role in thymus development
***Fgf8* hypomorphic allele**	Foregut endoderm from E8.0 then all 4 PPs, non-NC mesenchyme and surface ectoderm	Failure in formation of 3rd and 4th PPs		Required for PP formation Later role in thymus development, required for NC cell survival

Gene knockout/hypomorphic allele	Gene expression	Phenotype	Expression of other genes in transgenic	Possible role of gene in thymus organogenesis
***Fgfr2-IIIb*$^{-/-}$**	TECs, time of onset between E12.0 and E13.5	E12.5 TEC hypoplasia due to decreased proliferation	*Pax1* and *Pax9* normal at E13.5	Required for responding to Fgf7 and Fgf10 signals from mesenchyme
***RARα*$^{-/-}$ *RARβ*$^{-/-}$ and *RALDH2*$^{-/-}$**		Variable penetrance: failure in formation 3rd/4th PPs to thymic hypoplasia	Down-regulation of *Fgf8, Pax9, Tbx1*	Endoderm defect
***Hoxa3*$^{-/-}$**	3rd PA NC derived mesenchyme from E9.5, 3rd PP endoderm E9.5	Athymia (3rd PP forms but apoptotic from E11.5)	*Pax1* down-regulated only in the 3rd PP at E10.5 Loss of *Gcm2*	Rudiment proliferation and survival
***Hoxa3*$^{+/-}$ *Hoxb3*$^{-/-}$ *Hoxd3*$^{-/-}$**	*Hoxb3* and *Hoxd3* expressed in mesenchyme	Ectopic thymi and parathyroid glands		Caudal migration

Gene knockout/hypomorphic allele	Gene expression	Phenotype	Expression of other genes in transgenic	Possible role of gene in thymus organogenesis
***Pbx1*$^{-/-}$**	All four PPs, surrounding mesenchyme and ectoderm	Variable – athymia/ectopia/hypoplasia 3rd PP formation delayed	*Pax1* expression delayed *Gcm2* and *Tbx1* expression initiated at correct time may have lower expression level	Proliferation of rudiment
***Pax1*$^{-/-}$**	Foregut endoderm E8.0 All 4 PPs Adult thymus, restricted to a subpopulation of TEC	Hypoplastic/cystic thymi Delayed T cell maturation Decreased T cell number	MHC class II expression decreased (number of cells and expression level)	TEC development
***Pax9*$^{-/-}$**	Foregut endoderm E8.0 All 4 PPs Adult thymus	3rd/4th PP retarded from E11.5, ectopic/cystic thymi Few T cells, increased T cell death		Required for T cell survival

Gene knockout/hypomorphic allele	Gene expression	Phenotype	Expression of other genes in transgenic	Possible role of gene in thymus organogenesis
Eya1$^{-/-}$	1st to 4th PA 2nd to 4th PC 2nd to 4th PPs from E9.5	3rd pouch lost after E10.5 Athymic	No change in *Hoxa3, Pax1, Pax9* Loss of *Gcm2, Foxn1* Down-regulation of *Six1*	3rd pouch survival, apoptosis in PC
Gcm2$^{-/-}$	Dorsal portion of 3rd PP from E9.5	Aparathyroid No thymus phenotype		Parathyroid gland formation
Foxn1$^{-/-}$	Ventral portion of 3rd PP from E11.25	Athymic		TEC proliferation and differentiation
Shh$^{-/-}$	Foregut endoderm, excluded from 3rd PP until E10.5 when expressed at dorsal tip	Aparathyroid	No *Gcm2* *Hoxa3* normal *Foxn1* expressed throughout the primordia	Establishing parathyroid domain within the common primordia

1.3.1. Tbx1 is required for formation of the 3rd pharyngeal pouch

Mouse models of the human condition DiGeorge syndrome have uncovered a regulator of 3rd pharyngeal pouch formation, *Tbx1*. DiGeorge or Velocardiofacial syndrome is present in 1:4000 live births (Scambler 2003). Patients present with defects in tissues derived from the pharyngeal region including the thymus, parathyroids, facial dysmorphogenesis and cleft palate, as well as multiple cardiovascular abnormalities, which are a major cause of mortality. All aspects of the syndrome display highly variable penetrance, even between monozygotic twins. DiGeorge syndrome is caused by haploinsuffiency of the T-box containing gene, *Tbx1,* which is present on human chromosome 22 (Baldini 2005). Up to 29 other genes are also commonly included in the deleted region.

Tbx1 is expressed in the mouse pharyngeal endoderm from E8.0 and is strongly expressed at the time of formation of each of the four pharyngeal pouches (Chapman, Garvey et al. 1996; Vitelli, Morishima et al. 2002). Expression is also seen in non-neural crest derived pharyngeal arch mesenchyme and in the surface ectoderm surrounding the pharyngeal region. Mutant mice have been generated that contain a 1Mb targeted deletion in mouse chromosome 16 (referred to as *dfl*), which contains 18 of the 30 homologous genes deleted in the human syndrome including *Tbx1* (Lindsay, Vitelli et al. 2001). Additionally, a *Tbx1* null allele has been constructed by deletion of eight of the nine exons containing the *Tbx1* coding sequence, including deletion of the region encoding the T-box (Jerome and Papaioannou 2001).

Mice heterozygous for both the *dfl* and *Tbx1* null alleles recapitulate the cardiovascular aspects of DiGeorge. In $dfl^{+/-}$ mice, ectopic and/or hypoplastic thymi, a charateristic of the human condition, are seen in C57BL/6 and 129SvEv mice, but not on mixed backgrounds, indicating that genetic modifiers in addition to *Tbx1* haploinsufficiency are necessary to generate this phenotype (Taddei, Morishima et al. 2001). Both *Tbx1* null and $dfl^{-/-}$ mice die at birth, and display abnormalities in tissues formed and patterned from the pharyngeal region. These include

cardiovascular defects, which are the cause of death; smaller, shortened necks; abnormal face structure and ears; and usually complete absence of thymic or parathyroid tissue (although one lobe was found in a single *Tbx1*$^{-/-}$ embryo).

No defects in the formation of the 1st pharyngeal arch and pouch are seen in *Tbx1*$^{-/-}$ embryos. The more caudal arches and pouches are however severely hypoplastic. In keeping with this, both *Gcm2* and *Pax9,* which are expressed in the 3rd pouch endoderm (see sections 1.3.5 and 1.3.7), are strongly down-regulated in the pharyngeal region of *Tbx1*$^{-/-}$ embryos E9.5 compared to wild-type littermates ,as determined by global transcriptional analysis (Ivins, Lammerts van Beuren et al. 2005). *Pax1* is expressed in all four pharyngeal pouches (see section 1.3.7), and two analyses have looked at the expression of *Pax1* RNA in *Tbx1*$^{-/-}$ embryos at E9.5: 1) Microarray analysis of pharyngeal region tissues - including the foregut endoderm, surrounding mesenchyme and surface ectoderm - detected no change in *Pax1* expression (Ivins, Lammerts van Beuren et al. 2005), however 2) *in situ* hybridisation showed specific lack of *Pax1* expression in the 2nd to 4th pharyngeal pouches at E9.5 (Xu, Cerrato et al. 2005). Given the severe hypoplasia of the 2nd and 3rd pharyngeal pouches and arches in *Tbx1*$^{-/-}$ embryos, comparison of gene expression levels between *Tbx1* null and wild-type embryos by microarray analysis of intact pharyngeal region requires validation by *in situ* hybridisation or immunohistochemistry. It is therefore likely that *Pax1* expression is down-regulated in the 2nd and 3rd pharyngeal pouches and that the microarray result is due to expression of *Pax1* in the 1st pouch and the loss of tissue from the more caudal pharyngeal region. Expression of *Hoxa3* and *Eya1,* which are required for survival, patterning and proliferation of the 3rd pharyngeal pouch, were also unchanged in the microarray analysis, although again validation by *in situ* hybridisation or immunohistochemistry is required.

Analysis of mice in which *Tbx1* deletion was induced via use of the Cre-LoxP system (Cre^{ER}) has indicated that *Tbx1* is required at the time of formation of the 3rd pharyngeal pouch (Xu, Cerrato et al. 2005). Deletion of *Tbx1* at or before E9.5 results in athymia, whereas deletion of *Tbx1* after E10.5 results in thymic hypoplasia.

Thus, *Tbx1* is likely to have both early and late roles in thymus organogenesis. The distinct temporal requirement for *Tbx1* during development of the pharyngeal region, and the absence of caudal arch and pouch formation in null animals, has led to the suggestion that this gene may be required for segmentation. Additionally *Tbx1* may be required for proliferation of the pharyngeal endoderm, since a 50% reduction in cell number is seen at E10 in the absence of *Tbx1* (Xu, Cerrato et al. 2005). Specific down-regulation of Fgf8 (see next section) is seen in the endoderm of $Tbx^{-/-}$ embryos suggesting that Fgf8 is downstream of *Tbx1*. These gene products may also interact genetically, since $Tbx^{+/-}Fgf8^{+/-}$ mice have a more severe thymus phenotype than $Tbx1^{+/-}$ animals (Vitelli, Taddei et al. 2002). However, due to the expression of *Tbx1* in all three germ layers of the pharyngeal region and the abnormal neural crest migration seen in null embryos, the exact role that *Tbx1* plays in formation of the 3rd pharyngeal pouch and thymus organogenesis remains elusive.

1.3.2. Fgfs have multiple roles in thymus organogenesis

1.3.2.1. Fgf8 is essential for 3rd pharyngeal pouch formation

Fibroblast growth factors (Fgfs) are a family of secreted molecules that signal via cell surface receptor tyrosine kinases and have many roles in the developing embryo. In the pharyngeal region, Fgf8 is expressed in the foregut endoderm, surrounding non-neural crest mesenchyme and surface ectoderm from E8.0, and expression is later found in the pharyngeal pouches (Crossley and Martin 1995). Although analysis of Fgf8 function in thymus organogenesis in null embryos has not been possible due to early embryonic lethality, hypomorphic alleles in both mice (Abu-Issa, Smyth et al. 2002) and zebrafish (Reifers, Walsh et al. 2000) have been constructed. Fgf8 hypomorphic embryos ($Fgf8^{neo/-}$) have multiple neural, limb, craniofacial and cardiac abnormalities. The 1st and 2nd pouches are intact but fused (Abu-Issa, Smyth et al. 2002) however, the 3rd and 4th pharyngeal pouches and their corresponding arches fail to form, suggesting an early role for Fgf8 in 3rd and 4th pouch formation. Fgf8 is believed to signal through the receptor FGFR1. However unlike $Fgf8^{neo/-}$ embryos,

mice hypomorphic for this receptor have 2nd pharyngeal arch and pouch defects, whereas arches 1 and 3 are normal in appearance; unfortunately the specific 3rd pharyngeal pouch phenotype was not analysed in this study (Trokovic, Trokovic et al. 2003). Analysis of Fgf8 conditional mutants has revealed that, similar to *Tbx1,* removal of Fgf8 from the 3rd pouch endoderm and ectoderm at E9.5, after initial pouch formation, results in hypoplasia and ectopia of the thymus and parathyroid glands (Macatee, Hammond et al. 2003), indicating a continued requirement for Fgf8 in organogenesis. This defect may be due to a role for Fgf8 signalling between the 3rd pharyngeal pouch endoderm and the surrounding neural-crest derived mesenchyme, as increased apoptosis of neural-crest cells is seen in the hypomorphic mice (Abu-Issa, Smyth et al. 2002). Notably, experimental ablation of neural crest cells also results in thymic hypoplasia (Bockman and Kirby 1984; Veitch, Begbie et al. 1999).

1.3.2.2. Fgfr2-IIIb null animals have hypoplastic thymi

Initiation of *Fgfr2-IIIb* expression in the thymus occurs after E12, and by E13.5 expression is found throughout the epithelial compartment of the thymus (Revest, Suniara et al. 2001). The expression of two Fgfr2-IIIb ligands, Fgf7 and Fgf10, in the mesenchyme adjacent to the thymic primordia suggests that Fgfr2-IIIb could be involved in epithelial-mesenchyme interactions during thymus organogenesis (Revest, Suniara et al. 2001). In line with this observation, Fgfr2-IIIb and Fgf10 null embryos both have hypoplastic thymi due, at least in part, to a decrease in TEC proliferation (Ohuchi, Hori et al. 2000; Revest, Suniara et al. 2001). However, the *Fgfr2-IIIb*$^{-/-}$ thymus is detectably smaller than that of *Fgf10*$^{-/-}$ animals, suggesting that Fgf7 and Fgf10 may have overlapping functions during thymus organogenesis, although no thymus phenotype was reported in *Fgf7*$^{-/-}$ animals (Guo, Degenstein et al. 1996). No defects in initial formation of the primordia were reported in *Fgfr2-IIIb*$^{-/-}$ embryos, and both *Pax1* and *Pax9* were expressed at wild-type levels at E13.5. In addition parathyroid development was normal in *Fgfr2-IIIb*$^{-/-}$ embryos.

1.3.3. Retinoid signalling is required for 3rd pharyngeal pouch formation

The active derivative of vitamin A is retinoic acid (RA), which regulates gene transcription by binding to a retinoic acid receptor (RAR)/retinoid X receptor (RXR) heterodimer and plays multiple roles during embryo development (Chambon 1996). In the pharyngeal region, retinoid signalling is necessary for formation of the 3rd and 4th pharyngeal pouches. Analysis of cultured E8.0 embryos treated with a pan-RAR antagonist (Wendling, Dennefeld et al. 2000), mice hypomorphic for the retinaldehyde dehydrogenase 2 (RALDH2) enzyme responsible for synthesising RA in the pharyngeal region (Vermot, Niederreither et al. 2003), and embryos doubly null for RARα and RARβ (Dupe, Ghyselinck et al. 1999), all reveal variably penetrant defects ranging from failure to initiate the formation of the 3rd and 4th pouches and their corresponding arches to hypoplasia of these structures. These phenotypes are due to an intrinsic defect in the endoderm rather than to neural crest, as neural crest migration occurs normally (Dupe, Ghyselinck et al. 1999) and pouch formation is known to occur in the absence of neural crest (Veitch, Begbie et al. 1999).

RA is known to directly regulate expression of some *Hox* genes, via RA response elements (Chambon 1996). Expression of *Hoxa1* and *Hoxb1* is reduced in the pharyngeal region when RA signalling is limiting, suggesting that this mechanism may also operate in the pharyngeal region (Wendling, Dennefeld et al. 2000; Vermot, Niederreither et al. 2003). However, the pharyngeal pouch phenotype of *Hoxa1*$^{-/-}$ *Hoxb1*$^{-/-}$ embryos appears less severe than that described for RA deficiency (Rossel and Capecchi 1999), suggesting that RA also regulates expression of additional genes. Down-regulation of *Fgf8, Pax9* and *Tbx1* was also reported in the posterior region of the foregut endoderm upon removal of RA (Wendling, Dennefeld et al. 2000; Vermot, Niederreither et al. 2003), leading to the suggestion that only an enlarged 2nd pharyngeal pouch is formed. However due to the abnormal morphology of the region these results are difficult to interpret.

1.3.4. *Hoxa3* is required for formation of the thymic primordia

Hox genes are transcriptional regulators that contain a homeodomain DNA-binding motif, and play roles in segmentation and specification. Elegant genetic studies have given insight into the role the *Hox3* members, *Hoxa3, Hoxb3* and *Hoxd3* in development of the pharyngeal region. All three paralogues are expressed in neural crest cells that populate the third pharyngeal arch at E9.5. However *Hoxa3* alone is expressed in the third pharyngeal pouch endoderm (Manley and Capecchi 1998). In line with the expression patterns, *Hoxd3*$^{-/-}$ embryos have transformations of the cervical vertebrae (Condie and Capecchi 1993), while *Hoxb3*$^{-/-}$ mice are healthy and fertile but present with defects in the cervical vertebrae and IXth cranial nerve at low-penetrance (Manley and Capecchi 1998). *Hoxa3* null mice die shortly after birth due to circulatory problems, and exhibit a wide spectrum of defects including athymia; aparathyroidism; thyroid and tracheal abnormalities cervical vertebrae, cartilage and heart defects; and lack of the carotid body and IXth cranial nerve (Chisaka and Capecchi 1991; Manley and Capecchi 1995; Manley and Capecchi 1997; Kameda, Nishimaki et al. 2002).

The third pharyngeal pouch forms correctly in *Hoxa3*$^{-/-}$ embryos (Manley and Capecchi 1998; Kameda, Arai et al. 2004). However, although *Pax1* expression, which is required for proper thymus organogenesis (see 1.3.5), is initiated normally at E9.5, by E10.5 expression is specifically down-regulated in the third pharyngeal pouch (Manley and Capecchi 1998). The thymic and parathyroid regions of the pouch fail to extend at E11.5 and chromogranin A expression, which marks the parathyroid region from E11.5, was never detected in the null embryos (Kameda, Arai et al. 2004). Furthermore, increased apoptosis, observed in the primordium at E11.5, results in the disappearance of the structure by E12.5 (Kameda, Arai et al. 2004; Chisaka and Kameda 2005). Neural crest migration is unaffected in *Hoxa3*$^{-/-}$ embryos (Manley and Capecchi 1995; Chisaka and Kameda 2005), *Hoxa3* expression in the endoderm is normal in the absence of *Hoxa3*-expressing neural crest (Manley and Capecchi 1995), and *Hoxa3*$^{-/-}$ mice have tracheal defects which do

not require neural crest interactions, suggesting that the early failure in formation of the thymic and parathyroid rudiment may be due to an intrinsic defect in the endoderm. Alternatively it may be a result of loss of survival signals from the surrounding neural crest derived mesenchyme (Shinohara and Honjo 1997).

Hoxa3 heterozygous mice do not display overt thymic or parathyroid defects although they do have decreased serum levels of parathyroid hormone, suggesting haploinsufficiency (Su, Ellis et al. 2001). Although *Hoxb3* and *Hoxd3* single and double null embryos do not exhibit thymic or parathyroid defects, the additional loss of one functional copy of *Hoxa3* (*Hoxa3*$^{+/-}$ *Hoxb3*$^{-/-}$ and *Hoxa3*$^{+/-}$ *Hoxd3*$^{-/-}$) results in ectopic parathyroid glands and occasional ectopic thymi, due to a failure in caudal migration (Manley and Capecchi 1998). Heterozygosity for *Hoxa3* on a *Hoxb3* and *Hoxd3* double null background (*Hoxa3*$^{+/-}$ *Hoxb3*$^{-/-}$ *Hoxd3*$^{-/-}$) results in a fully penetrant failure of thymus and parathyroid migration. These results suggested that there is a quantitative requirement for *Hox* gene expression in thymus and parathyroid organogenesis. Additionally, *Hoxa3* may be required at multiple stages during development of these organs, including an early role in rudiment proliferation and a later role in primordia migration. Since *Hoxa3* is expressed in the thymus at all stages of embryonic development and in the adult, it may also have a role in the mature organ (Manley and Capecchi 1995; Manley and Capecchi 1998), however conditional analysis is required to test this hypothesis.

1.3.5. *Pbx1* null mice have delayed 3rd pharyngeal pouch formation and thymic defects

Subsets of Hox genes, including the Hox3 paralogues, are believed to mediate transcriptional activation after binding DNA in combination with members of the *Pbx* family of homeobox-containing genes (Jacobs, Schnabel et al. 1999; Ferretti, Marshall et al. 2000). Expression of *Pbx1* is found in the third pharyngeal pouch endoderm, surrounding mesenchyme and surface ectoderm (Manley, Selleri et al. 2004) and *Pbx1* null mice display a wide spectum of defects in tissues derived from

all three germ layers, which result in embryonic lethality by E16 (Selleri, Depew et al. 2001). Derivatives of the third pharyngeal pouch, the thymus and parathyroid, display variable defects including bilateral absence, ectopic and hypoplastic primordia (Manley, Selleri et al. 2004). The hypoplastic nature of the rudiments results, at least in part, from decreased proliferation. Unlike *Hoxa3* null embryos, formation of the third pharyngeal pouch is affected in *Pbx1*$^{-/-}$ embryos, being both delayed and abnormal. *Pax1* expression was detected, but did not appear until E10.5 and then only in a very small domain, whereas in *Hoxa3* null mice *Pax1* expression is initiated normally but is then down-regulated from E10.5 (Manley and Capecchi 1995). In contrast, *Gcm2,* which marks the dorsal, parathyroid region of the pouch, was initiated at the correct time although the level of expression was lower than in wild-type. Expression of *Tbx1* in the dorsal third pouch was similarly down-regulated.

1.3.6. *Pax1* and *Pax9* are expressed throughout thymus organogenesis

Members of the paired box (*Pax*) family of transcription factors play essential roles in regional specification during early organogenesis in several organ systems (Mansouri, Goudreau et al. 1999). *Pax1* and *Pax9* are highly homologous members of this family, and contain a DNA binding-paired domain, an octopeptide sequence commonly found in *Pax* genes. However, they do not contain the paired-like homeodomain that is present in many other family members (Gruss and Walther 1992; Wallin, Mizutani et al. 1993). *Pax1* and *Pax9* are expressed in overlapping but non-identical patterns in the embryonic and adult thymus and also in the sclerotome, facial mesenchyme and limb buds (Deutsch, Dressler et al. 1988; Neubuser, Koseki et al. 1995; Muller, Ebensperger et al. 1996; Wallin, Eibel et al. 1996). In the pharyngeal region expression of both *Pax1* and *Pax9* is initiated in the ventral floor of the foregut endoderm at E8.5 and subsequently expression is found throughout the four pharyngeal pouches (Deutsch, Dressler et al. 1988; Neubuser, Koseki et al. 1995). Studies using mutant mice suggest that *Pax1* (Dietrich and Gruss 1995;

Wallin, Eibel et al. 1996) and *Pax9* (Hetzer-Egger, Schorpp et al. 2002) both have roles regulating development of the thymic and parathyroid primordia, although neither appears to be essential for the initial formation and patterning of the primordia from pharyngeal endoderm.

1.3.6.1. Pax1 null mice have a hypoplastic thymus and defects in T cell development

Evaluation of the role of *Pax1* in development has been aided by the identification of a number of naturally occurring *Pax1* mutant mice, known as undulated or *un* mutants. Un mice were initially collected and bred for their short kinky tails, which are caused by irregularly shaped vertebrae and intervertebral discs (Dietrich and Gruss 1995). Four *Pax1* mutants have been preserved which display varying severity of phenotype (Dietrich and Gruss 1995; Adham, Gille et al. 2005) and an additional *Pax1* null line has been generated by gene targeting (Adham, Gille et al. 2005). All of these mutants have hypoplastic, cystic thymi, which contain reduced numbers of T cells, in addition to sclerotome defects.

Two hypomorphic *Pax1* alleles have been described, the original *un* mutant, which contains a point mutation in the paired domain (Wallin, Eibel et al. 1996), and the recently identified *un* scoliosis (un^{sco}) caused by deletion of exons 1 to 4 and the 5' region flanking the gene (Adham, Gille et al. 2005). Analysis of both *un/un* and un^{sco}/un^{sco} embryos at E17.5 reveals a delay in T cell maturation, evidenced by an increase in the number of T cells negative for both CD4 and CD8 (DN) and a proportional decrease in double positive (DP) cells. The *un* extensive (*ex*) mutant contains a point mutation in the paired domain that results in a null allele (Wallin, Eibel et al. 1996), and these mice show a two-fold reduction T cell number and an increase in the DP/DN ratio at E17.5, as reported for the *sco* mutant. A fourth, naturally occurring mutation of *Pax1*, short-tail (*s*), is semi-dominant in heterozygous mice and homozygous lethal (Dietrich and Gruss 1995). In this mutant, a large deletion event has completely removed the *Pax1* gene, and the semi-dominant phenotype results from ectopic expression of an adjacent gene, the transcription

factor *Nkx2-2,* in the *Pax1* expression domains of the somites and limb buds (Kokubu, Wilm et al. 2003). Although no ectopic *Nkx2-2* expression was reported in the pharyngeal region of un^{s} mice, heterozgotes display a two-fold decrease in T cell number at E17.5 (Wallin, Eibel et al. 1996), suggesting that either there is an undetected level of mis-expression or that further, as yet unidentified, genes may be affected.

In addition to expression in the developing pharyngeal region, *Pax1* is expressed in a subset of MHCII negative, cortical TEC in the adult thymus (Dietrich and Gruss 1995; Wallin, Eibel et al. 1996), suggesting that it may have a further role in maintenance of the adult thymus, indeed in adult un^{sco}/un^{sco} mice T cells numbers remain reduced although the DN:DP ratio is normal.

1.3.6.2. Pax9 null mice have cystic and ectopic thymi and severe defects in T cell production

$Pax9^{-/-}$ neonates die just after birth, probably due to a cleft palate (Peters, Neubuser et al. 1998), and display a more severe thymus phenotype than *Pax1* null mice (Hetzer-Egger, Schorpp et al. 2002). Development of $Pax9^{-/-}$ thymi appears normal until E11.5, when development of both the 3rd and 4th pharyngeal pouches appears to be retarded. When formed, thymi are ectopically located in the larynx, due to a failure in caudo-ventral migration. Additionally, they are severely hypoplastic, contain cysts and display very compromised thymopoiesis. The $Pax9^{-/-}$ thymus is colonised by T cell precursors, however few mature α/β T cells are produced and increased levels of apoptosis are seen as development proceeds, suggesting that critical survival signals are not provided by TEC. No γ/δ T cells are present due to an apparent lack of T cell receptor γ-chain expression, as determined by *in situ* hybridisation. IL-7 is currently the only factor known to be absolutely required for production of γ/δ T cells (Laky, Lefrancois et al. 2000), however, IL-7 expression was normal in $Pax9^{-/-}$ mice. Thus, *Pax9* could be required independently of IL-7 for γ/δ T cell production, or the authors failed to detect either a subtle change in IL-7 expression or the presence of a very small γ/δ T cell population. Interestingly, despite

the broad expression of *Pax1* and *Pax9* in the pharyngeal pouches, only defects in the thymus, derived from the ventral portion of the 3rd pharyngeal pouch, have been described in *Pax1*$^{-/-}$ and *Pax9*$^{-/-}$ mice. Generation of embryos doubly mutant for *Pax1* and *Pax9* may illuminate roles for these genes in other pharyngeal pouch derivatives that were previously undetected due to functional redundancy. Furthermore, it is possible that a more severe defect in thymus organogenesis may be detected.

1.3.6.3. Maintenance of Pax1 and Pax9 expression in the 3rd pharyngeal pouch requires Hoxa3

Pax1 and *Pax9* are specifically down-regulated in *Hoxa3* null mice at E10.5, although their initial expression is unaffected, suggesting that *Hoxa3* is upstream of *Pax* genes in thymus organogenesis (Manley and Capecchi 1995). Certainly, there is genetic interaction between *Hoxa3* and *Pax1,* as double heterozygotes (*Hoxa3*$^{+/-}$ *Pax1*$^{+/-}$) display a thymus phenotype (hypoplastic thymi and thymopoietic defects) that is indistinguishable from *Pax1* null mice (Su and Manley 2000; Su, Ellis et al. 2001). Loss of an additional copy of *Pax1* results in a more severe phenotype; *Hoxa3*$^{+/-}$*Pax1*$^{-/-}$ have ectopic thymi containing cysts and severe defects in thymopoiesis – specifically, decreased levels of immature $CD4^+$ $CD8^+$ thymocytes as a result of increased levels of apoptosis by their immediate precursors - caused by TEC malfunction (Xu, Woo et al. 1997; Hetzer-Egger, Schorpp et al. 2002). The T cell defect of *Hoxa3*$^{+/-}$ *Pax1*$^{-/-}$ null animals may be due to decreased MHC class II expression by TECs, as MHC class II is required for progression from double negative to double positive stages of T cell maturation (Anderson, Jenkinson et al. 1993). Thus, *Pax1* and *Pax9* are downstream of *Hoxa3* in a common signal transduction pathway.

1.3.7. Expression of *Eya1* and *Six1* are necessary for formation of the common thymus/parathyroid primordia

Eya genes contain a conserved protein-protein interaction domain (Eya) and a divergent transactivation domain. They play roles in organ and sensory development

in vertebrates, probably by controlling the expression of inductive signals (Xu, Woo et al. 1997). *Eya1* is expressed in the 1st to 4th pharyngeal arches, the 2nd to 4th pouch endoderm and the surface ectoderm including the 2nd to 4th cleft regions from E9.5 (Xu, Zheng et al. 2002). All four pharyngeal pouches form in *Eya1*$^{-/-}$ embryos. However, the 3rd pharyngeal pouch is lost after E10.5 as neither derivative is detected. Abnormalities in the ultimobranchial bodies derived from the 4th pharyngeal pouch are also present (Xu, Zheng et al. 2002). *Eya1* may be downstream of, or in a different pathway from, the *Hox* and *Pax* genes since *Hoxa3*, *Pax1* and *Pax9* expression at E9.5 and E10.5 are unaffected in *Eya1* null mice (Xu, Zheng et al. 2002).

Six family members are down-stream of *Eya* genes in a pathway regulating eye development, which is conserved from *Drosophila* to mouse (Halder, Callaerts et al. 1995). *Six1*$^{-/-}$ mice are also athymic (Laclef, Souil et al. 2003) and expression of *Six1* was down-regulated in the endoderm of *Eya1*$^{-/-}$ mice, suggesting that this pathway operates during in thymus development. The parathyroid specific gene, *Gcm2* is not expressed in *Eya1* null mice suggesting that parathyroid organogenesis is not initiated. Since no thymus organogenesis appears to occur (no *Foxn1* expression was detected), despite a lack of alteration in *Hox* and *Pax* gene expression in the pharyngeal region prior to E10.5, after which the pouch can no longer be detected, it is likely that either *Eya1* plays a direct role in initiation of the joint parathyroid/thymus primordium or that other, unidentified genes are affected. Although no change in levels of cell death were seen in the arch mesenchyme or pouch endoderm at E10.5, a significant increase in apoptosis was present in the surface ectoderm. *Eya1* is therefore required for ectoderm cell survival. Thus, the loss of ectoderm may affect endoderm survival, or the endoderm defect in *Eya1*$^{-/-}$ embryos may be due to an intrinsic, or mesenchymal defect.

1.3.8. *Gcm2* is required for formation of the parathyroid gland

Gcm2 is one of two mammalian homologues of the *Drosophila* gene *glials cells missing* (*Gcm*), which is necessary and sufficient for the formation of glia in the developing nervous system. *Gcm* family members have a gcm-motif, which mediates base specific DNA binding. Unlike *Drosophila gcm*, neither of the mouse homologues are expressed in the developing nervous system. *Gcm1* is expressed in the labyrinthine trophoblast cells of the placenta and *Gcm1*$^{-/-}$ embryos exhibit placental failure. *Gcm2* expression marks the dorsal region of the 3rd pharyngeal pouch from E9.5 when the pouch forms (Gordon, Bennett et al. 2001), and subsequently marks the parathyroid region of the common thymus/parathyroid primordia where it is co-expressed with parathyroid hormone (PTH). Although *Gcm2*$^{-/-}$ embryos display a complete lack of parathyroid glands, suggesting that *Gcm2* specifies the parathyroid, the immediate fate of dorsal 3rd pouch cells that should express *Gcm2* have not been described in *Gcm2* null animals. *Gcm2*$^{-/-}$ mice are viable and fertile and have normal serum PTH levels (Gunther, Chen et al. 2000) and, intriguingly the source of PTH in these mice has been suggested to be a small cluster of cells expressing *Gcm1* present under the capsule of the thymus. This same group of cells is also found in wild-type mice. Since *Gcm2* expression was down-regulated in *Hoxa3*$^{-/-}$, *Hoxa3* may regulate *Gcm2* expression in the dorsal 3rd pharyngeal pouch. However, the function and down-stream pathways of *Gcm2* have not been investigated.

1.3.9. *Foxn1* is required for TEC differentiation and recruitment of T cell precursors

Mutations in the transcription factor *Foxn1,* a member of the forkhead family of proteins, result in athymia and hairlessness as seen in the classical mouse mutant *nude* (Flanagan 1966; Cordier and Haumont 1980; Nehls, Pfeifer et al. 1994). *Foxn1* is not required for initiation of thymus organogenesis (Nehls, Pfeifer et al. 1994), but is required cell autonomously for further TEC development (Blackburn, Augustine et

al. 1996). In keeping with this, *Foxn1* is expressed at a high level throughout the thymic primordium from E11.25, after patterning of the 3rd pharyngeal pouch and formation of the thymic primordium has occurred (Gordon, Bennett et al. 2001). However, expression can be detected by RT-PCR in the 3rd pharyngeal pouch at E10.5 (Balciunaite, Keller et al. 2002) and in whole E9.0 embryos (Nehls, Pfeifer et al. 1994). *Nude* mice have a thymic rudiment, which undergoes correct migration to the ventral mid-line, but is alymphoid and highly cystic (Cordier and Haumont 1980). Thymic epithelial cells of the postnatal *nude* thymus display extensive staining for the thymic epithelial progenitor cell markers, MTS20 and MTS24, and lack defined cortical and medullary regions, suggesting that they have failed undergo proliferation and differentiation (Blackburn, Augustine et al. 1996; Nehls, Kyewski et al. 1996). This result is consistent with the role of *FoxnI* in the skin, where it is responsible for initiation of keratinocyte terminal differentiation (Baxter and Brissette 2002; Janes, Ofstad et al. 2004) and were ectopic *Foxn1* induces keratinocyte hyperproliferation (Prowse, Lee et al. 1999).

Further insight into the role *Foxn1* plays in TECs has been gained from a hypomorphic *FoxnI* allele ($FoxnI^{\Delta/\Delta}$) generated by insertion of a GFP transgene into exon three, which results in Foxn1 protein lacking the domain encoded by exon three due to an in frame exon two – exon four splice (Su, Navarre et al. 2003). Like the null thymic rudiment, defined cortical and medullary regions cannot be identified in the postnatal $FoxnI^{\Delta/\Delta}$ thymus, and most epithelial cells have an immature, $K8^+$ $K5^+$ phenotype. However, in contrast to *nude* mice, the thymi of $FoxnI^{\Delta/\Delta}$ contain T cells, although their numbers are 2-5% of wild-type numbers and a majority are immature, double negative cells. Therefore, either expression of *Foxn1*, including the exon three domain, is required in TECs for T-cell – epithelial cell interactions that are necessary for the proliferation and differentiation of both cell types, or the expression level of *Foxn1* achieved in this mutant is insufficient for the cross-talk to occur. This work also suggests that the requirement for *Foxn1* (either expression level or exon three domain) is different at different sites of expression, as no skin or hair defects were identified.

Although *Foxn1* is required cell-autonomously for TEC differentiation, it does not specify the thymus. This has been established by three lines of evidence; 1) *nude* mice contain a correctly specified thymic rudiment (Cordier and Haumont 1980), 2) the *nude* thymic rudiment expresses IL-7, a TEC-specific marker required for T cell development (Zamisch, Moore-Scott et al. 2005), additionally *Foxn1*$^{-/-}$ cells can correctly initiate expression of a reporter gene present in the *Foxn1* locus (Nehls, Kyewski et al. 1996; Gordon, Bennett et al. 2001), and 3) the region of the foregut endoderm that is going to give rise to the 2nd and 3rd pharyngeal pouches, is capable of forming a functional thymus when grafted under the kidney capsule of nude mice (Gordon, Wilson et al. 2004).

1.3.10. *Wnts* can regulate *Foxn1* expression *in vitro*

Wnt proteins are secreted glycoproteins that play a diverse range of roles during development, including cell fate specification, proliferation, migration and cell death. Multiple *Wnt*s are expressed in the thymus, both in the epithelial and lymphoid compartments (Balciunaite, Keller et al. 2002). *In vitro* data from cortical and medullary cell lines suggests that the canonical/β-catenin Wnt signalling pathway is active in the thymus, and that Wnt4 and 5b can directly up regulate *FoxnI* expression (Balciunaite, Keller et al. 2002). The expression domains of Wnt4 and 5b overlap with that of *FoxnI,* suggesting that they may regulate *FoxnI* expression *in vivo*. However, this has not as yet been demonstrated, nor has this possibility been eliminated, as construction of a double *Wnt4* and *Wnt5b* knockout has not been reported.

1.3.11. Opposing gradients of *Shh* and *Bmp4* may regulate patterning of the 3rd pharyngeal pouch endoderm

Opposing concentration gradients of sonic hedgehog (Shh) and bone morphogenetic proteins (Bmps) are used during development to provide positional information to cells present in a single tissue (Sasai and De Robertis 1997). Both families of molecules have multiple roles in thymus organogenesis. Shh expression is excluded from the 3rd pharyngeal pouch until E10.5, at which point expression is seen in the most dorsal portion of the 3rd pharyngeal pouch, where it meets the foregut endoderm (Moore-Scott and Manley 2005). In *Shh*$^{-/-}$ embryos *Foxn1* expression is found throughout the 3rd pharyngeal pouch and into the pharynx, while no *Gcm2* expression can be detected. As *Hoxa3* expression is unchanged in these embryos either *Hoxa3* and Shh may independently regulate expression of *Gcm2* in the 3rd pharyngeal pouch or *Hoxa3* regulation of *Gcm2* may be Shh-dependent.

In contrast to Shh, Bmp4 expression is restricted to the thymic domain of the 3rd pouch at E10.5 and is also expressed in the 1st and 2nd pharyngeal arch mesenchyme and the dorsal portion of the 2nd pouch (Moore-Scott and Manley 2005). These expression and transgenic analyses suggest that gradients of Shh and Bmp4 pattern the 3rd pharyngeal pouch. Preliminary ectopic expression analyses using beads soaked with either Bmp4 or the Bmp antagonist noggin support this hypothesis, since ectopic Bmp4 leads to expansion of the *Foxn1*-expression domain whereas ectopic noggin results in a small expansion of the *Gcm2* domain (J. Gordon, UGA, Athens, personal communication). Noggin expression is found in the dorsal region of the pouch at E10.5 (J. Gordon, PhD thesis). However, Bmp4 may not be required at least for maintenance of *Foxn1* expression since expression of *noggin* under the control of a *Foxn1* promoter construct does not markedly affect expression of *Foxn1* at early stages of thymus development (Bleul and Boehm 2005), although this analysis did not include quantitative analysis of *Foxn1* expression. These mice exhibit ectopic, hypoplasic and cystic thymi, without defects in T cell development, although T cell number is drastically reduced. Genetic deletion of the Bmp receptor, BmpR1A using

a *Foxn1-Cre* also results in formation of an ectopic and hypoplastic thymus (J. Gordon, UGA, Athens, personal communication), consistent with the *Foxn1-noggin* transgenic results.

Bmps may also play a role in formation of the posterior pharyngeal pouches. Chordin (chrd) is a Bmp antagonist and is expressed in the dorsal region of the foregut endoderm at E8.5 (Bachiller, Klingensmith et al. 2003). *Chrd*$^{-/-}$ embryos form a single anterior pharyngeal pouch and arch, and *Tbx1, Fgf8* and *Pax1* are specifically down-regulated in *Chrd*$^{-/-}$ pharyngeal endoderm (Bachiller, Klingensmith et al. 2003). As ectopic chordin can induce both *Tbx1* and *Fgf8* expression in Xenopus embryos, chordin may regulate the transcription of these genes in the pharyngeal endoderm.

1.3.12. Summary

The data discussed indicate that RA, *Tbx1,* Fgf8, and the Bmp antagonist Chordin are all required by the pharyngeal endoderm to segment the pharyngeal region and form the pharyngeal pouches, including the 3rd pharyngeal pouch. Chordin is required for expression of both *Tbx1* and *Fgf8*, and *Fgf8* is specifically down-regulated in the endoderm in the absence of *Tbx1,* as determined by *in situ* hybridisation. RA may also regulate *Tbx1* and *Fgf8* expression, since expression of these genes is also lost when RA signalling is inhibited. However, due to the loss of the structures of the pharyngeal region, interpreting expression results is complex.

Hoxa3 has been placed at the head of a transcription factor network that is responsible for patterning of the 3rd and 4th pharyngeal pouches. *Hoxa3* is not responsible for induction of *Pax1* and *Pax9* expression in the 3rd pharyngeal pouch, as expression of both genes is initiated normally in the absence of *Hoxa3*. However, it is required to maintain expression in the 3rd pharyngeal pouch from E10.5, although it is unclear whether this is an endoderm-autonomous effect. *Hoxb3* and *Hoxd3* expression is present in the neural crest-derived mesenchyme surrounding the

3rd pharyngeal pouch, and although these genes are not necessary for formation of the thymic primordia they may play a later role in its chordo-ventral migration. Neither *Pax1* nor *Pax9* are necessary for defining the thymic primordium, but both genes are required for normal TEC development and therefore for normal thymopoiesis. Lack of either gene may result in a specific TEC deficiency, as thymopoiesis is arrested at defined points in development and increased thymocyte apoptosis is seen.

Hox genes regulate transcription in combination with members of the *Pbx* family. *Pbx1* is expressed in the pharyngeal region. In contrast to *Hoxa3*$^{-/-}$ embryos, 3rd pouch formation is delayed in the absence of *Pbx1*. However, the mice go on to form thymic and parathyroid organs, although they are commonly ectopic, hypoplastic and cystic. Thus, either *Hoxa3* and *Pbx1* independently regulate gene transcription in the 3rd pharyngeal pouch, or they both have additional regulatory partners.
Eya1 is required for production of the common thymic and parathyroid primordium. However, as with *Hoxa3,* it is difficult to establish the precise cause of the pharyngeal defect in *Eya1*$^{-/-}$ mice, as *Eya1* is expressed in tissues descending from all three germ layers. Since *Hoxa3, Pax1* and *Pax9* expression is unaffected until E10.5 in *Eya1* null mice *Eya1* maybe part of a different signalling pathway. *Eya* genes are down-stream of *Pax* family members in both eye and muscle development (Bonini, Leiserson et al. 1993; Cheyette, Green et al. 1994; Halder, Callaerts et al. 1998; Heanue, Reshef et al. 1999). *Six* genes are also present in this transcription factor pathway and *Six1* was down-regulated in the endoderm of *Eya1*$^{-/-}$ mice, suggesting that this pathway may be operating during thymus development. However, demonstration of this interaction may require the generation of *Pax1*$^{-/-}$ *Pax9*$^{-/-}$ as the thymic phenotype of single null mice is mild as compared to *Eya1*$^{-/-}$.

Gcm2 expression, which marks the dorsal region of the 3rd pharyngeal pouch from E9.5, is affected in many of the mutant mice discussed, including those proposed to be present in independent pathways. *Hoxa3, Pbx1, Eya1* and *Shh* null mice all form a 3rd pharyngeal pouch which is morphologically normal until at least E11.5 (Manley and Capecchi 1998; Xu, Zheng et al. 2002; Manley, Selleri et al. 2004; Moore-Scott

and Manley 2005). *Gcm2* expression is never detected in *Hoxa3*$^{-/-}$, *Eya1*$^{-/-}$ or *Shh*$^{-/-}$, suggesting that there is a 3rd pouch patterning defect in the absence of all three genes. However, *Hoxa3, Eya1* and *Shh* may regulate different aspects of 3rd pouch patterning, as there is currently no indication of genetic interaction between them. Expression of *Gcm2* is correctly initiated in *Pbx1* null animals but the level of expression may be lower than wild-type, indicating that patterning of the pouch has occurred but further defects are present. Unlike *Gcm2* expression, *Pax1* expression was delayed in the *Pbx1* null animals, this result suggests that *Pbx1* involves different pathways in the regulation of *Gcm2* and *Pax1* in the 3rd pouch. Additionally, *Pbx1* may regulate early *Pax1* expression in the 3rd pharyngeal pouch, whereas *Hoxa3* may independently regulates *Pax1* from E10.5.

The transcription factor, *Foxn1,* is only expressed at levels detectable by *in situ* hybridisation once thymic organogenesis has been initiated. *Foxn1* is required for TEC development subsequent to primordium formation, and is thus necessary for the production of a functional thymus. While *Hoxa3, Pax1, Pax9* and *Eya1* obviously play important roles in thymus development, they also regulate development of the parathyroid gland; *Hoxa3*$^{-/-}$, *Pax1*$^{-/-}$, *Pax9*$^{-/-}$ and *Eya1*$^{-/-}$ mice all have defects in parathyroid gland formation (Dietrich and Gruss 1995; Manley and Capecchi 1995; Hetzer-Egger, Schorpp et al. 2002; Xu, Zheng et al. 2002). *Gcm2* expression is restricted to the parathyroid domain of the 3rd pharyngeal pouch two days before *Foxn1* comes on in the thymic domain (Gordon, Bennett et al. 2001). It is therefore likely that there are other factors involved in specification of the thymus and induction of *Foxn1* expression. Recent work has suggested that opposing gradients of Shh and Bmp2/4 may determine the extent of the thymus and parathyroid domains (Moore-Scott and Manley 2005), and whilst it is clear that in the absence of Shh the entire 3rd pouch forms thymus, the reciprocal experiment with Bmp2/4 has not been reported. However, deletion of the Bmp receptor, Bmpr1, or expression of noggin, in the *Foxn1* spatial and temporal expression domain using a *Foxn1-Cre* may not effect expression of *Foxn1.* Although, this result does not rule out an early role for Bmp2 or 4 in induction of *Foxn1,* is does suggest that neither are required for maintenance of expression.

The very complex interactions and morphogenetic cell movements that occur during development of the pharyngeal region and thymus organogenesis suggest that many genes and gene products are involved. Whilst factors have been identified that are necessary for these processes to occur, we currently have very little real understanding of how the 3rd pharyngeal pouch forms and is patterned and how it gives rise to a functional thymus.

1.4. *Rhox* and homeobox-containing genes

1.4.1. The *Rhox* family of homeobox containing genes

The homeodomain is a 60 amino acid motif which mediates base-specific binding to DNA via formation of a helix-turn-helix structure. The homeobox was initially identified as a highly conserved sequence present in *Drosophila* homeotic selector genes, which define the identity of the developing body segments (Lewis 1978). Orthologues of the *Drosophila* HOX genes have been found in an ever-increasing number of species, and their role in segmentation has been conserved from flies to humans (Krumlauf 1992; Ferrier and Minguillon 2003). Although initially identified in *Hox* genes, the homeobox motif has been recruited and used by a wide range of genes, which play a role in many cellular processes both during development and in the adult. A gene that is found to encode a homeodomain is therefore likely to bind DNA, but the presence of a homeodomain does not suggest a function. With the exception of a few key residues, homeodomain proteins are highly diverged. However, with a few exceptions, they can however be subdivided into classes based on homeodomain sequence and the presence of other motifs.

The mouse *Rhox* family of homeobox-containing genes is highly diverged from the original *Hox* genes (Maclean, Chen et al. 2005). There are twelve *Rhox* family members, present in the A2 region of the X chromosome (Figure 1.4). Family members are defined by their divergent homeodomain, which is encoded by three exons, their location on the X chromosome and their similar length. The twelve *Rhox* genes are located in three distinct clusters within a 0.7Mb interval; the α cluster contains *Rhox1* through to *Rhox4*, the β cluster *Rhox5* to *9* and γ cluster *Rhox10, 11* and *12*. *Rhox* family members are primarily expressed in reproductive and extra embryonic tissues, although three members are each expressed in one additional tissue - including *Rhox4,* which is expressed in the developing and adult thymus

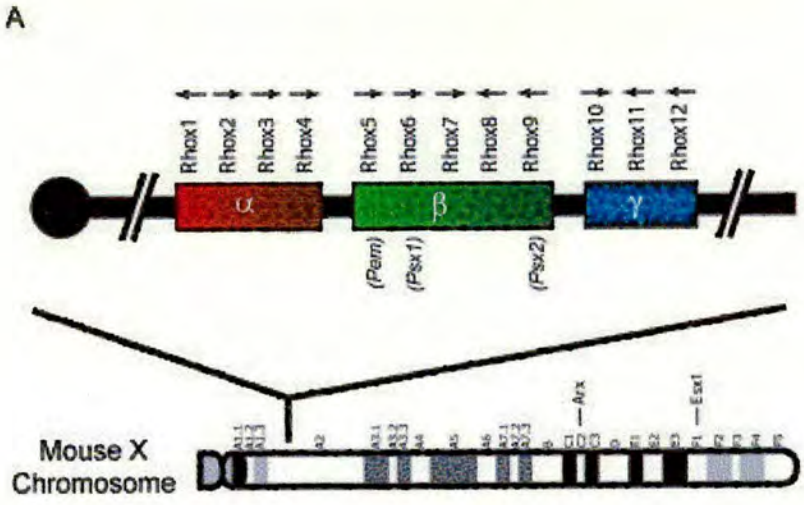


Figure 1.4 Diagram of the *Rhox* locus present in the A2 region of the mouse X chromosome. The twelve mouse *Rhox* genes are present in three gene clusters, α, β and γ. Gene orientation is indicated by the direction of the arrows. Diagram taken from (Maclean et al. 2005).

(Lin, Labosky et al. 1994; Han, Park et al. 1998; Chun, Han et al. 1999; Takasaki, McIsaac et al. 2000; Jackson, Baird et al. 2002; Jackson, Baird et al. 2003; Kang, Li et al. 2004; Maclean, Chen et al. 2005). Due to their relatively recent discovery, very little is know about the roles *Rhox* genes play in these tissues.However, knockout mice have been constructed for two family members. *Rhox5* null mice have a mild reduction in male fertility due to a defect in the rate of sperm production (Maclean, Chen et al. 2005), while *Rhox9*$^{-/-}$ have no discernable phenotype (Takasaki, Rankin et al. 2001), possibly due to the overlapping function of *Rhox* family members in reproductive and extra-embryonic tissues. Androgens regulate many aspects of male reproduction; it is therefore interesting that expression of at least five members of the mouse *Rhox* family, *Rhox2, 3, 5, 10* and *11* (Maclean, Chen et al. 2005) and one human *Rhox* gene (Geserick, Weiss et al. 2002) is upregulated by androgen treatment. Androgen responsiveness has not been assessed for any other mouse *Rhox* genes.

Although the *Rhox* family has only been described in mice (Maclean, Chen et al. 2005), a *Rhox5* rat orthologue (Sutton and Wilkinson 1997) and two human homologues (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002) also exist. Interestingly, rat *Rhox5* and the *Rhox5* gene from twelve sub-species of mice were found to have undergone rapid divergence and positive selection (Sutton and Wilkinson 1997), whereby selection events have occurred for nucleotide polymorphisms that result in a change in amino acid code. This feature is common among testis expressed genes (discussed below). The two human *Rhox* genes are present in the syntenic region of the human X chromosome (Xq24) (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002) and to date, no other human family members have been identified. Both human genes are highly diverged from their mouse counterparts, and human and mouse *Rhox* families may therefore share a single common ancestral gene (Maclean, Chen et al. 2005). Although expression of both human genes is also found in the testis, neither is expressed in the placenta, a tissue in which expression of all twelve mouse genes can be detected albeit at varying levels. Placenta formation and structure differs widely between mammals, possibly due to the recent evolution of this tissue and/or the ideal location of the

placenta for establishing barriers between different species (Hemberger 2002). Morphologically, mouse and human placentas are very similar. However, expression of the *Rhox* gene family in mouse but not human placenta suggests that molecular control of this tissue may differ. Mouse and human *Rhox* genes may however play a common role in the testis, since at least one of the human genes can also be regulated by androgens (Geserick, Weiss et al. 2002).

1.4.2. Co-linear expression of the *Rhox* family

Co-linear gene expression was first reported for the HOX clusters of *Drosophila* where it was observed that the chromosomal gene order corresponded with the order of gene expression down the anterior-posterior axis (Lewis 1978). This is now referred to as spatial co-linearity and has been observed in the *Hox* clusters of wide range of species including mammals. Two other co-linear mechanisms have been identified in mouse; temporal and quantitative co linearity, where the gene order corresponds to the time and level of gene activation respectively. It is becoming clear that co-linear expression of clustered genes is established and maintained by a range of different mechanisms, including gene transcriptional availability regulated by chromatin, signalling molecule concentration and global enhancers situated outside of the clustered genes (Kmita and Duboule 2003; Deschamps and van Nes 2005).

Each mouse *Rhox* cluster (α, β and γ) is reported to display quantitative co-linear expression during testis development. Additionally the α and γ clusters are reported to exhibit temporal co-linearity, whereby a different *Rhox* gene is activated at each stage of testis development (Maclean, Chen et al. 2005). A subtle difference exists between the co-linear expression seen within the *Rhox* cluster and that of the classical HOX clusters; the *Rhox* genes are turned on and then rapidly down-regulated as the next gene is up-regulated, so that one gene is active at one time, where as *Hox* genes are up-regulated and remain expressed as the next gene is being expressed, so that many *Hox* genes are expressed in one location at one time (Kmita and Duboule 2003). This difference may be due to different control mechanisms

controlling gene expression. Since there are only two *Rhox* genes in the human genome, it is unlikely that a similar observation will be made in the human testis and this may reflect differences in testis development between the two species.

1.4.3. *Rhox* genes and the X chromosome

The *Rhox* family are unique in being the only cluster of homeobox genes present on X, and are all expressed in reproductive and extra embryonic tissues, which are over-represented on the X chromosome as compared to autosomes (Hemberger 2002; Vallender and Lahn 2004; Vallender, Pearson et al. 2005). The reasons for this bias are much debated, but are likely to be due to phenomena experienced only by the sex chromosomes. These include the presence of a single X chromosome in males, resulting in hemizygous exposure and the possibility of fixing recessive male beneficial mutations in the population. Secondly, there are two X chromosomes in females, so a female-specific gene will be present twice as often in females as in males (sexual antagonism).

Duplicated genes on the *Drosophila* X chromosome tend to be more diverged than duplicates located on autosomes (Thornton and Long 2002), again probably due to sexual selection. Whether this observation is also true of mammalian X chromosomes is unclear. However, the mouse *Rhox* cluster probably arose through gene duplication and *Rhox* genes are highly diverged, showing as low as 30% amino acid identity in the homeodomain. This observation may additionally be due to the possible role of *Rhox* family members in male reproduction, since rapid divergence has now been reported for many testis genes, including *Rhox5* (Sutton and Wilkinson 1997; Ting, Tsaur et al. 1998; Wyckoff, Wang et al. 2000; Wang and Zhang 2004).

1.5. Gene duplication

1.5.1. Gene duplication

Current knowledge of gene and protein structure indicate that a limited number of protein and nucleotide motifs found in nature have been recruited to mediate differing functions in various cellular processes. Usually, at least one motif is found in each gene or protein, and the same motifs are found throughout evolution. The primary mechanism that has allowed, and continues to allow, these processes to occur is gene duplication (Ohno 1970).

Gene duplication events resulting from duplication of an entire genome, an individual chromosome, a chromosomal segment or stretch of genomic DNA have all occurred in evolution. Mechanisms mediating duplication events include; abnormal chromosome segregation, abnormal sister chromatid crossing over, DNA slippage and transposition via transposable elements (Figure 1.5) (Achaz, Netter et al. 2001). Duplicated genes present either at a location remote from, or inverted with respect to, their original gene copy are likely to have been generated initially via tandem duplication, and been inverted or redistributed later (Achaz, Netter et al. 2001). A classical example of a gene family having arisen through gene duplication is the *Hox* family. Here, an ancient duplication initially created a single *Hox* cluster, which in mammals has been duplicated twice, both completely (HoxA and HoxD) and incompletely (HoxB and probably HoxC) to create four clusters on four different chromosomes, each containing up to 13 paralogues (Akam 1989; Krumlauf 1992). Although generated by gene duplication, the *Hox* clusters of vertebrates (but not invertebrates) now appear to be protected from further events as tandem duplications have not been identified in mammals and transposable elements are excluded from the vicinity of these clusters (Wagner, Amemiya et al. 2003).

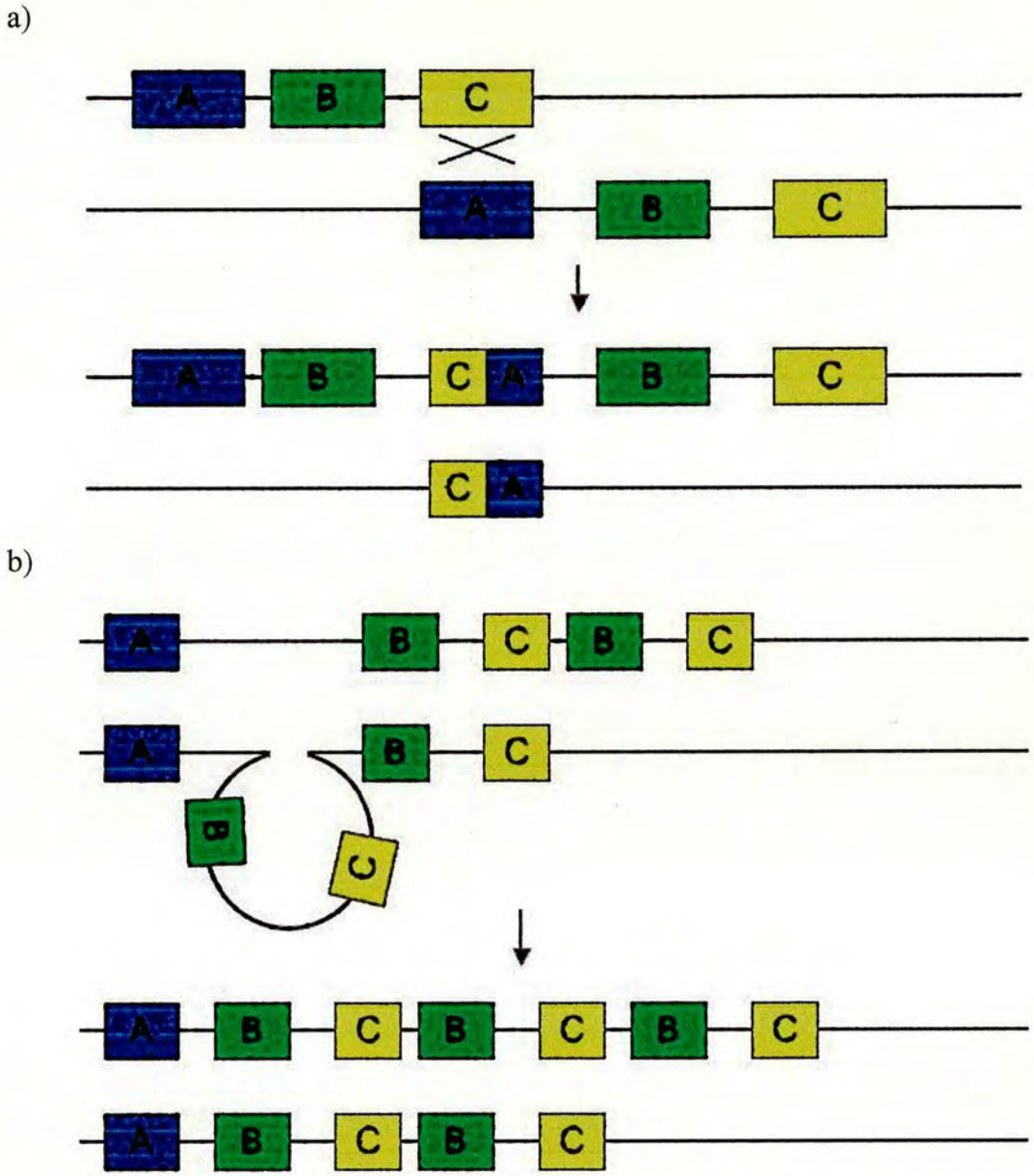


Figure 1.5 Two mechanisms of gene duplication. a) abnormal crossing over between sister chromatids during meiosis after asymmetric pairing, this can occur between homologous genomic sequences including homologous genes and transposable elements. b) DNA slippage results from mispairing of chromatids during replication and recombination creating loops of DNA, which generate duplications and deletions when repaired.

1.5.2 Fate of duplicated genes

The study of complete and partial genome sequences indicates that in mammals, around 1% of genes duplicate every one million years (Lynch and Conery 2000). Although up to 80% of gene duplicates are subsequently lost from the genome, a significant proportion (up to 50%) are maintained and become fixed in the population. Due to functional redundancy after a duplication event, gene duplicates experience a period of moderately relaxed selective pressure, which allows mutations to accumulate. Mathematical models of gene mutation rates predict that the amount of time required for a gene to become non-functional (to generate a pseudogene) is relatively short: around four million years (Force, Lynch et al. 1999). Recently, there has been much debate about the evolutionary processes that lead to preservation of duplicates. Classical studies suggest that duplicates are preserved only if they acquire mutations that give the gene a new function (neo-functionalisation) (Ohno 1970), and that this process may have played an important role in the evolution of new morphological structures (Force, Cresko et al. 2005). How often neo-functionalisation occurs is unknown, however, it is unlikely to be frequent enough to account for the large proportion of duplicates retained in most genomes.

Many genes are multi-functional, for example, being expressed in more than one tissue or at multiple times during development; having a role in a number of biological processes; being alternately spliced to form different protein products; and/or being composed of multiple functional domains. It has therefore been proposed that duplicates can undergo sub-functionalisation, whereby different gene copies retain complementary functions allowing the survival of both gene copies in the genome, and many examples appear to exist (Force, Lynch et al. 1999; Lynch and Force 2000; Makova and Li 2003; Aharoni, Gaidukov et al. 2005; Kopelman, Lancet et al. 2005). Sub-functionalisation may not require active selection but could be a neutral process (Force, Cresko et al. 2005). A number of factors influence whether sub-functionalisation is likely to occur. For instance, a small population size is required for both genes to be fixed in the population; the bigger and more complex a gene, the more likely it is that functions can be divided between duplicates.

Whether all regulatory elements and structural features are inherited by the duplicate is an additional and important factor (Lynch and Katju 2004; He and Zhang 2005). Subfunctionalisation and possibly neofunctionalisation has occurred in the *Hox3* cluster; *Hoxa3* is required for thymus organogenesis and uniquely for *Hox3* paralogues is expressed in the endoderm (Chisaka and Capecchi 1991; Manley and Capecchi 1995), while *Hoxd3* is essential for the correct specification of the cervical vertebrae, a process in which *Hoxa3* is not thought to function (Condie and Capecchi 1993).

1.5.3 Duplications present in the mouse genome

Over 80% of genes present in the human genome have a single mouse orthologue (Waterston, Lindblad-Toh et al. 2002), and contained within remaining 20% are species specific genes. Analysis of the mouse genome sequence has identified clusters of gene families which have undergone large expansions via gene duplication and are specific to the mouse. These families are primarily involved in reproduction, immunity and olfaction (Waterston, Lindblad-Toh et al. 2002; Emes, Goodstadt et al. 2003; Huminiecki and Wolfe 2004), all of which play a role in survival. It is no coincidence that genes from these families are additionally likely to have undergone rapid evolution (and sometimes positive selection), resulting in a high level of sequence polymorphism between orthologues from different species and also between different subspecies. The combination of enhanced rates of gene duplication and rapid evolution has resulted in the divergence of mouse reproductive mechanisms, creating species barriers and generating biodiversity (Lynch and Conery 2000; Taylor, Van de Peer et al. 2001; Emes, Goodstadt et al. 2003).

An example of such a species specific cluster is the mouse *Rhox* cluster (Maclean, Chen et al. 2005), members of which are expressed in reproductive tissues and at least one member of which has undergone positive selection (Sutton and Wilkinson 1997). There are twelve *Rhox* family members in mice and only two in human (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002; Maclean, Chen et al.

2005). An additional feature of this cluster, commonly observed amongst species specific duplications, is that the expression patterns of members from two different species are unlikely to be conserved (Huminiecki and Wolfe 2004); thus, the mouse *Rhox* genes are expressed in reproductive and extra-embryonic tissues whereas the human genes are only expressed in reproductive tissues (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002; Maclean, Chen et al. 2005).

1.5.4 Gene duplication and gene conversion

Gene conversion is a recombination mechanism that copies a DNA sequence from one stretch of genomic DNA to another (Radding 1978). Gene conversion events are thought to be common amongst gene families, especially when paralogues have retained a high degree of sequence identity (Ahn, Dornfeld et al. 1988; Elliott, Richardson et al. 1998) and remained in close physical proximity (Galtier 2003). They result in a high degree of sequence identity being maintained between gene duplicates, making old duplicates appear young and preventing elimination of gene copies from the genome. In transgenic systems the rate of gene conversion between duplicates is so high that it has been suggested that mechanisms may exist to inhibit these events, otherwise the efficiency of gene conversion would prevent tandem gene duplicates from diverging (Murti, Bumbulis et al. 1992). However, gene conversion events are not uniform between all gene duplicates, one bias that has been identified is for DNA sequences with a high GC content. Therefore, regions with high gene content are favoured by gene conversion, and further gene conversion events can result in increased GC content by favouring GC basepairs (Galtier 2003; Marais 2003). However, despite the potential significance of gene conversion mechanisms for gene evolution, very little is known about the role gene conversion has played, and continues to play in sculpting genomes.

1.6 Aims

The data discussed in 1.3 indicate that no genes have yet been identified that are responsible for specification of the thymus. *Foxn1* is expressed specifically in the thymic domain of the 3rd pharyngeal pouch and is required for TEC differentiation, however the thymus is already specified by the time of onset of *Foxn1* expression. All other regulators of thymus development identified to date also affect parathyroid development. Several possible models could reconcile these data into specification mechanisms, including the presence of an as yet unidentified thymus factor. Previous work carried out in this lab showed that *Rhox4* is expressed in the thymic domain of the 3rd pharyngeal pouch endoderm at E9.5, and might therefore play a role in specification of this organ.

In this thesis, I set out to investigate whether *Rhox4* might play an important role in early thymus organogenesis, and in particular, whether it is responsible for specification of the thymus. Chapter 3 describes expression of *Rhox4* during thymus organogenesis, and shows that the expression pattern is consistent with a role in specification. However, I was unable to map *Rhox4* into known transcription factor pathways. In Chapter 4 I describe characterisation of the *Rhox4* locus, a prerequisite to generation of null and conditional-null alleles for functional analysis. This analysis uncovered tandem duplication events in the *Rhox4* locus, containing seven *Rhox4* copies. Additionally, the evolutionary history of these duplications is addressed. The identification of seven *Rhox4* copies necessitated further analysis of copy expression patterns which is described in Chapter 5. Finally, Chapter 6 describes the design and construction of a transgenic RNAi strategy for establishing the role of *Rhox4* in thymus organogenesis.

Identification of genes expressed in anterior endoderm during ES cell differentiation

Afifah Rahman, Anestis Tsakiridis, Richard Axton, Lesley Forrester and Josh Brickman.

Institute for Stem Cell Research, The University of Edinburgh, Roger Land Building, King's Builings, West Mains Road, Edinburgh, EH9 3JQ, UK.

The precise nature of signals and mechanisms underlying the process of anterior-posterior patterning in mammals is not entirely understood. One of the earliest molecular asymmetries is seen by the expression of homeobox gene Hex, which is expressed asymmetrically in the peri-implanting embryo. We have an ES cell line that reports reliably on Hex expression during ES cell differentiation. This line contains a dsRed in exon 1 of the Hex locus. We will use a GFP labelled gene trap vector to identify genes co-expressed with Hex during ES cell differentiation. Our gene trap vectors are unique in that they include a splice acceptor, followed by β-galactosidase and hygromycin resistance to monitor expression and splice donor-neomycin to select for integration into class II genes. We will present data on the functionality of these gene traps alongside data from a pilot screen for co-expression with Hex. The nature of the gene trap will enable the identification of both direct and indirect targets. This approach is more economical than standard in vivo gene-trap screening because it is possible to determine if likely knock-out alleles are expressed in one or more specific embryonic tissues before generating chimaeric or transgenic mice.

Chapter 2 Materials and Methods

2.1 Materials

Unless otherwise stated, analytical grade chemicals were obtained from either Sigma, BDH Laboratory supplies or Fisher Scientific. Flowgen supplied electrophoresis grade agarose. Synthetic oligonucleotides were synthesised by Sigma.

2.2 Molecular biology

Basic molecular biology techniques were performed as described in *Molecular Cloning – A Laboratory Manual* (Sambrook, Fritsch et al. 1989).

2.2.1 Isolation of nucleic acids

2.2.1.1 Plasmid preparation

Small scale plasmid purification was performed using QIAprep miniprep spin kits (Quiagen). Large scale plasmid purification was performed using Plasmid Maxi Kits (Quiagen).

2.2.1.2 Genomic DNA isolation

DNA from fresh liver: Tissue was chopped into small pieces then incubated overnight at 55°C with 10ml DNA lysis buffer (100mM Tris, pH8.5; 5mM EDTA; 0.2% SDS; 200mM NaCl; 100μg/ml proteinase K). Undigested tissue was removed by centrifugation at 3000xg for 5 minutes. Supernatant was put in a fresh tube and 10ml of phenol:chloroform:isoamyl alcohol (25:24:1) added, vortexed and spun at 10000rpm for 10 minutes. The aqueous phase was then transferred to a fresh tube and the DNA was precipitated after adding 2 volumes of isopropanol. Samples were mixed by rocking, the DNA precipitate was allowed to settle and the aqueous phase

was poured off. The DNA pellet was washed in 5ml 70% ethanol, air dried at room temperature and left to dissolve overnight at 4°C in 1 ml sterile water. DNA from inbreed mouse strains: C57BL/6, 129, CBA and wild-type *Mus musculus castaneous* mice as well as DA rat was prepared in this manner. Genomic DNA from other wild mouse strains mentioned was purchased from Jackson Labs.

Ear punch DNA: 100µl of DNA lysis buffer was added to ear punches in eppendorf tubes and incubated overnight at 55°C. 100µl of isopropanol was then added and tubes were shaken vigorously by hand to precipitate the DNA. Tubes were then spun for 10 minutes at 10000rpm in a microcentrifuge. Supernatants were removed and the DNA pellets were washed in 70% ethanol. DNA was then resuspended in 100µl of sterile water.

2.2.1.3 Isolation of RNA

50-100 mg of freshly isolated tissues were homogenised in 1ml of Trizol reagent (Gibco BRL). RNA was isolated from cells by incubating 5-10x10^6 cells in 1ml of Trizol reagent. Homogenisation was performed at 15-30°C. 0.2ml of chloroform was then added to homogenised samples in eppendorf tubes and then shaken vigorously by hand for 15 seconds. Samples were incubated at 15-30°C for 2-3 minutes and centrifuged at 12000xg for 15 minutes at 2-8°C. Aqueous phases were transferred to fresh eppendorf tubes and RNA was precipitated by adding 0.5ml isopropanol and mixing the samples. Following this, samples were incubated at 15-30°C for 10 minutes and centrifuged at 12000xg for 10 minutes at 2-8°C. supernatants were removed and RNA pellets were washed in 1ml 75% ethanol. Samples were mixed by gentle vortexing and then spun at 7500xg for 5 minutes at 2-8°C. RNA pellets were then air dried for 10 minutes and resuspended in RNAse free water. To ensure the RNA was fully dissolved, samples were incubated at 55°C for 10 minutes. Human testis RNA was purchased from Clontech.

2.2.2 Molecular cloning

2.2.2.1 Conventional DNA cloning

DNA plasmids were digested with the appropriate restriction enzymes and fragments were separated by gel electrophoresis on a 0.8% agarose gel. Fragments were then isolated using Quiagen gel extraction kits. The plasmid backbone was dephosphorylated by incubation with 1U of Antarctic Phosphatase (NEB) for 15 minutes at 37°C, followed by inactivation at 75°C for 10 minutes. Purified plasmids and inserts were ligated at 1:1 and 1:3 DNA ratios using Quick ligation kit (NEB). Reaction mixtures were incubated for 2 hours at 37°C and transformed into chemically competent *E. coli*. Colonies containing the ligated products were isolated by plating on agar plates containing the appropriate antibiotic selection.

2.2.2.2 Cloning of oligonucleotides

The sequences of oligonucleotides used for introducing the human H1 promoter, a polylinker and the shRNA sequences are shown below. Each oligonucleotide was synthesised and 5' phosphorylated by Sigma. Oligonucleotide pairs (called forward and reverse below) were then annealed by incubating 5µl of a 100µM solution of complementary oligonucleotides at 70°C for 10 minutes, then removing them to a beaker of water at 37°C and allowing them to cool to room temperature. 1µl of a 1:1000 dilution of the annealed oligonucleotides were then used in each cloning reaction

Oligonucleotide sequences:

Human H1 promoter forward:

5'AATTCTTAATTAAAATATTTGCATGTCGCTATGTGTTCTGGGAAATCAC
CATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGA
GACCACTCGGTACCAT3'

Human H1 promoter reverse:

5'GAATTAATTTTATAAACGTACAGCGATACACAAGACCCTTTAGTGGTAT
TTGCACTTTACAGAAACCTAAACCCTTAGAATATTCAAGACATACTCTGG
TGAGCCATGGTAGATC3'

Polylinker ligated 3' of the human H1 promoter forward:

5'CTAGAGGATCCGGCGCGCCCCGC3'

Polylinker ligated 3' of the human H1 promoter reverse:

5'TCCTAGGCCGCGCGGGG3'

RNAi1 forward:

5'GTACCAACACCAACTACCTACTTCATTTCAAGAGAAAATGAAGTAGGT
AGTTGGTGTTTTTTTGGAAAT3'

RNAi1 reverse:

5'CTAGATTTCCAAAAAAACACCAACTACCTACTTCATTTTCTCTTGAAAT
GAAGTAGGTAGTTGGTGTTG3'

RNAi2 forward:

5'GTACCAAGCTGCCAGAGGAGCCAGTTTTCAAGAGAAAAACTGGCTCCT
CTGGCAGCTTTTTTTGGAAAT3'

RNAi2 reverse:

5'CTAGATTTCCAAAAAAAGCTGCCAGAGGAGCCAGTTTTTCTCTTGAAAA
CTGGCTCCTCTGGCAGCTTG3'

RNAi4 forward:

5'GTACCGGGACAAAGCAGAAGAATTAATTCAAGAGAAATTAATTCTTCT
GCTTTGTCCCTTTTTGGAAAT3'

RNAi4 reverse:
5'CTAGATTTCCAAAAAGGGACAAAGCAGAAGAATTAATTTCTCTTGAAT
TAATTCTTCTGCTTTGTCCCG3'

RNAi5 forward:
5'GTACCTTTAAGAAGAGGAGAGAGCACTTCAAGAGAAAGTGCTCTCTCC
TCTTCTTAAATTTTTGGAAAT3'

RNAi5 reverse:
5'CTAGATTTCCAAAAATTTAAGAAGAGGAGAGAGCACTTTCTCTTGAAG
TGCTCTCTCCTCTTCTTAAAG3'

AvrII containing RNAi1 forward:
5'GTACCAACACCAACTACCTACTTCATTTCCTAGGAAATGAAGTAGGTA
GTTGGTGTTTTTTTGGAAAT3'

AvrII containing RNAi1 reverse:
5'CTAGATTTCCAAAAAAACACCAACTACCTACTTCATTTCCTAGGAAATG
AAGTAGGTAGTTGGTGTTG3'

AvrII containing RNAi2 forward:
5'GTACCAAGCTGCCAGAGGAGCCAGTTTTCCTAGGAAAACTGGCTCCTC
TGGCAGCTTTTTTTGGAAAT3'

AvrII containing RNAi2 reverse:
5'CTAGATTTCCAAAAAAAGCTGCCAGAGGAGCCAGTTTTCCTAGGAAAA
CTGGCTCCTCTGGCAGCTTG3'

AvrII containing RNAi4 forward:
5'GTACCGGGACAAAGCAGAAGAATTAATTCCTAGGAATTAATTCTTCTG
CTTTGTCCCTTTTTGGAAAT3'

AvrII containing RNAi4 reverse:
5'CTAGATTTCCAAAAAGGGACAAAGCAGAAGAATTAATTCCTAGGAATT
AATTCTTCTGCTTTGTCCCG3'

AvrII containing RNAi5 forward:
5'GTACCTTTAAGAAGAGGAGAGAGCACTTCCTAGGAAGTGCTCTCTCCT
CTTCTTAAATTTTTGGAAAT3'

AvriII containing RNAi5 reverse:
5'CTAGATTTCCAAAAATTTAAGAAGAGGAGAGAGCACTTCCTAGGAAGT
GCTCTCTCCTCTTCTTAAAG3'

2.2.2.3 PCR cloning

The specificity of PCR products was confirmed by gel electrophoresis and aliquots taken directly from the PCR reaction mixtures were used for each cloning reaction. PCR products containing a polyA overhang were cloned using the pCR2.1-TOPO plasmid (Invitrogen) using the manufacturers instructions. Products that gave a blunt product were cloned using the pCR-Blunt II-TOPO plasmid and the Zero Blunt TOPO PCR Cloning Kit (Invitrogen).

2.2.2.4 Chemically competent *E. coli*

E.coli (DH5α) were grown overnight in 5ml LB medium (1% w/v tryptone; 0.5% w/v yeast extract; 85mM NaCl), this starter culture was then transferred to 500ml LB medium and grown until a cell density reading of between A_{600}=0.3-0.6 was reached to ensure the cells were in log phase. Cells were then chilled by swirling on ice and centrifuged at 3000rpm in pre-chilled bottles. Cells were resuspended in 1/20 volume ice cold, sterile LB pH6.1, supplemented with 10% PEC Mw 3350; 5% DMSO; 10mM $MgCl_2$; 10mM $MgSO_4$ and 10% glycerol, incubated on ice for 10 minutes. Cells were aliquoted into eppendorf tubes on dry ice, then stored at -80°C.

2.2.2.5 Transformation of chemically competent *E. coli*

20µl of DNA ligation mixture was added to 20µl of KCM (0.5M KCl; 0.15M $CaCl_2$; 0.25M $MgCl_2$) to a final volume of 100µl with sterile water. The resulting mixture was then chilled on ice for 5-10 minutes. 100µl of KCM competent *E. coli* was then added and incubated on ice for 20 minutes. Following this, the mixture was incubated at room temperature for 10 minutes. 1ml of LB medium was then added and incubated at 37°C for 60 minutes with shaking at 200rpm. Mixtures were spun at 12000rpm for 30 seconds and resuspended in 100µl of LB medium. Cells were then plated onto agar plates (1.5% w/v agar in LB) containing the appropriate antibiotic selection.

2.2.3 PCR

PCR reactions in this thesis were carried out using three *Taq* DNA Polymerase enzymes, *Taq* DNA Polymerase (Roche) was used for all PCR reactions where the product was not intended for sequencing. Two proofreading enzymes were used for producing products for cloning, *PfuTurbo* Hotstart DNA Polymerase (Stratagene) and Platinum *Pfx* DNA Polymerase (Invitrogen). Details of reaction mixtures used for each enzyme are:

Taq DNA Polymerase (Roche)

1µl cDNA/DNA
1x buffer
0.1mM dNTPs
0.7µM each primer
0.6U enzyme

Platinum *Pfx* DNA Polymerase (Invitrogen):

1µl cDNA

1x buffer

1mM $MgSO_4$

0.3mM dNTPs

0.3µM each primer

0.5U enzyme

PfuTurbo Hotstart DNA Polymerase (Stratagene):

1µl cDNA

1x buffer

3mM $MgSO_4$

0.3mM dNTPs

0.3µM each primer

1U enzyme

2.2.4 Primers used for all PCR and RT-PCR reactions

Details of primers used for all PCR and RT-PCR reactions are shown in Table 2.1. All primers were synthesised by Sigma.

Table 2.1 Details of primers used in this thesis.

Gene name	Primer name	Sequence	Annealing temperature (°C)	cDNA product size	Figure
mRhox4	5060	TCAGGACATGGAGCATCAAA	64.3	708	3.5, 3.6, 5.2,
	Ehox exon 4	TGCAATAACATGCTGGTGGAAG	66.8		5.7, 5.9, 5.10
mRhox4	Ehox forward	TGGAGCATCAAAACACCAAC	66.2	336	5.1, 5.8, 6.2,
	Ehox reverse	GGAATAGGCTGCACTTTGTC	68.3		6.3
mRhox4	Ehox exon 2F	CACTGGATGGAGAGGGAAGA	64.2	369	4.1, 4.7
	Ehox exon 2R	CGGATGAAGTGATTCTGCTG	63.4		
mRhox4	GeneRacer 5'	CGACTGGAGCACGAGGACACTGA	74		5.4
	5' RACE rev2	AGCAGGCCCACCTTCTCCACTTAA	71.4		
mRhox4	GeneRacer 3'	GCTGTCAACGATACGCTACGTAACG	72		5.5
	3' RACE fwd1	GGAGGCCTGGTGTACCACCCTTGA	74.5		
rRhox4	Rat Ehox fwd	GCCTCGTGGATAACAGGAAC	63.4	450	5.11
	Rat Ehox rev	CTTGAGCAAGGGTGGTTCTC	63.7		
rRhox4	Rat EE2 fwd	GTGCAAAGGCACAGAAGGTT	64.1	429	4.18
	Rat EE2 rev	CCGATGAAGTGATTCTGCTG	63.4		

Gene name	Primer name	Sequence	Annealing temperature (°C)	cDNA product size	Figure
β-actin	Bact 5/6F	CGTTGACATCCGTAAAGACC	60	206	All RT-PCR experiments
	Bact 5/6R	ATCTGCTGGAAGGTGGACAG	62		
β-actin	GeneRacer 5'	CGACTGGAGCACGAGGACACTGA	74	872, 828	5.6
	Control Primer B1	GACCTGGCCGTCAGGCAGCTCG	72		
β-actin	GeneRacer 3'	GCTGTCAACGATACGCTACGTAACG	72	1800, 1715	5.6
	Control Primer A	GCTCACCATGGATGATGATATCGC	76		
mRhox1	Rhox1 fwd	GAGGAAAATGCGAATGGTGT	63.7	344	5.8, 5.9, 5.10
	Rhox1 E2 rev	GATGTACTGAGTGCGCTGGA	64.1		
mRhox2	Rhox2 fwd	GCGACAAAGCGTCAATTACA	63.7	473	4.9, 5.8, 5.9, 5.10
	493053I12	GCCTCCTTAGTGCTGATGTAGTG	63.8		
mRhox3	Rhox3 fwd	GAAGATGGAGGACAGGTGGA	64.1	344	4.10, 5.8, 5.9, 5.10
	Tex2 rev	GGCCAGTTGTTTTCTTGCTT	63		
rRhox3	Rat XM343759R	TGGCCTTGAAAATGCTCTCT	63.7	326	5.11
	Rat XM343759RF	TGCTGACTGAAGAGGGAAGAA	64		
mRhox5	Rhox5 fwd	GCAAGGTCACCAGGCTACTC	63.8	321	5.8
	Rhox5 rev	CTACCCCCAGGATTTCCATT	63.6		

Gene name	Primer name	Sequence	Annealing temperature (°C)	cDNA product size	Figure
mRhox6	Rhox6 fwd	TCAAGAAGAGCCTGCTCCAT	64	386	5.8
	Rhox6 rev	ATCCGAAACCAATTCTGCAC	63.7		
mRhox7	Rhox7 fwd	AAGCATGGATGGGTCTGAAG	64	364	5.8
	Rhox7 rev	TACACTCAGGGACCCGATTC	63.8		
mRhox8	Rhox8 fwd	GAAGAATGGAGCCAATGGAA	63.8	736	5.8
	Rhox8 rev	GGTAAACCTGTAGCGGTTGC	63.3		
mRhox9	Rhox9 fwd	TCAAGAAGAGCCTGCTCCAT	64	483	5.8
	Rhox9 rev	GCTGCTCCAAAATCTTCAGG	63.7		
mRhox10	Rhox10 fwd	ATGTCGCTTTGGAAGAGGTG	64.1	370	5.8
	Rhox10 rev	TCGAGGTTCCAGATGTAGCA	63.4		
mRhox11	Rhox11 fwd	GGGATGAGCTTCTGTGAGGA	64.4	322	5.8
	Rhox11 rev	TTTCTTTTGAGGGCATCTGG	63.9		
mRhox12	Rhox12 fwd	CAAACTGGGGGAGAATGAGA	63.9	398	5.8
	Rhox12 rev	TGCTTTGCAAGGTTCTTCCT	63.7		
mFoxn1	Foxn1 E3 Fwd	CTCCAGAGAGGACACCCTCA	64.5	367	3.8
	Foxn1 E3 Rev	CTCTGCTGGGAAGCTAGGC	64.1		

Gene name	Primer name	Sequence	Annealing temperature (°C)	cDNA product size	Figure
hPEPP1	hPEPP1 F2	CACCGTGTTCTACTGCCTGA	63.9	350	5.12
	hPEPP1 R2	CGGAAAACACTTTCCAGCTC	63.6		
hPEPP1	hPEPP2 F2	GCAGTGCAGATTTGGTTTGA	63.8	332	5.12
	hPEPP2 R2	TTGGGGAATGTGAAAGAAGG	63.7		
hHoxa3	hHoxa3 fwd	CTGGATGAAAGAGTCTCGAC	60	379	5.13
	hHoxa3 rev	CCAGCGAATGCATAGAGTTC	60		
hPaxI	hPax1 fwd	CCAACGTGGTCAAGCACATC	62	512	5.13
	hPax1 rev	CTAAGGCAGGTTTCTCTAGC	60		
hFoxnI	hFoxn1	GAAACCTGTGGGAACAGTTG	60	365	5.13
	hFoxn1	ACTTCCAGACCAGGCAAACG	62		

2.2.4.1 cDNA synthesis

First strand cDNA was synthesised from 5μg of total RNA using oligo dT with Superscript II Reverse Transcriptase (Invitrogen) following the manufacturers instructions. Following cDNA synthesis, DNase treatment was performed using the DNase treatment kit (Ambion). 1μl cDNA was then used in each PCR reaction.

2.2.5 RT-PCR

2.2.5.1 5' and 3' RACE

5' and 3' RACE reactions were performed using GeneRacer Kit (Invitrogen). Briefly, full length cDNA was synthesised from 5μg of RNA from both a mix of RNA from all embryonic tissues (whole E9.5 embryo, whole E15.5 embryo, E15.5 placenta, E11.5 yolk sack and ES cell line) and adult tissues (brain, heart, lung, muscle, skin, testis and thymus) known to express *Rhox4* (see Figure 5.1) and HeLa cell line RNA which was provided in the kit as a control. 5' and 3' RT-PCR reactions were then carried out using a gene specific primer and 5' or 3' RACE primer (see Table 2.1), with Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen).

PCR reaction mixture for 5' and 3' RACE (see Table 2.1 for details of generacer and gene specific primers):

Generacer primer (10μM)	3μl
Gene specific primer (10μM)	1μl
cDNA	1μl
Buffer	5μl
dNTPs (10mM)	1μl
Taq (1U/μl)	0.5μl
$MgSO_4$ (50mM)	2μl
dH_2O	36.5μl

PCR cycle parameters 5' and 3' RACE (touch down):

Cycle name	Step	Temperature (°C)	Time (seconds)	Number of cycles
Initial denaturation	Denaturation	94	120	1
72°C	Denaturation	94	30	5
	Annealing	72	45	
70°C	Denaturation	94	30	5
	Annealing	70	45	
63°C	Denaturation	94	30	30
	Annealing	63	30	
	Extension	72	45	
Final extension	Extension	72	600	1

PCR products were gel purified using QIAquick Gel Extraction Kit (Qiagen) and shotgun cloned using Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen).

2.2.5.2 Real-time RT-PCR

cDNA synthesis was performed as described in section 1.2.4. cDNA was diluted five-fold with sterile water and 5µl was added to the reaction mixture described below:

2µl 5 x Lightcycler-DNA Master SYBR Green reaction mixture (Roche) (containing *Taq* DNA polymerase, reaction buffer, dNTP, SYBR green I dye and 1mM $MgCl_2$)

0.1µM forward primer

0.1µM reverse primer

sterile water to 10µl

Each reaction was loaded into capillary tubes, centrifuged at 3000rpm for 30 seconds and loaded into the Lightcycler (Roche) and run according to the programme shown below:

Denaturation:	95°C for 30 seconds
Amplification:	95°C for 10 seconds, 55°C for 5 seconds, 72°C for 20 seconds, 85°C for 2 seconds
Melting curve:	65°C – 95°C at 0.1°C/second
Cooling:	95°C – 40°C

Fluorescence was acquired at the end of each amplification cycle and throughout the melting curve analysis. To enable the derivation of a standard curve for quantification, dilution series containing 10^6, 10^5, 10^4 and 10^3 copies of each gene analysed were also performed using plasmids containing clones of each PCR generated product.

2.2.6 DNA sequencing

DNA sequencing was performed on plasmid DNA with 3.2pmol of sequencing primer and was carried out by the School of Biological Sciences Sequencing Service using Big Dye Terminator reaction mix (ABI).

2.2.7 Southern blotting

2.2.7.1 Restriction enzyme digestion of genomic DNA

80U of the appropriate restriction enzyme was added to 10μg of genomic DNA. The reaction mixtures were then incubated overnight at 37°C. Following this, a further 40U of restriction enzyme was added to the reaction mixture and incubated for 2-4 hours at 37°C.

2.2.7.2 Blotting

Restriction enzyme digested genomic DNA was run on a 0.8% agarose gel containing 0.5μg/ml ethidium bromide overnight. The gel was then photographed alongside a fluorescent ruler to allow verification of band sizes against a 1kb ladder. Following this, the gel was rinsed in dH_2O and washed for 20 minutes in depurination buffer (0.25M HCl). A second rinse in dH_2O was performed and then the gel was washed for 2x20 minutes in denaturation buffer (1.5M NaCl; 0.5M NaOH). A third rinse in dH_2O followed and then the gel was washed for 2x20 minutes in neutralisation buffer (1.5M NaCl; 0.5M Tris; pH7.0). the gel was then rinsed for a final time in dH_2O. To blot the DNA onto a membrane, the gel was placed upside down on a Whatman 3M paper wick soaked in 2xSSC, the ends of which were dipped in a basin containing 20xSSC. A Hybond N nylon membrane was pre-wetted in dH_2O and then placed on top of the gel. A small amount of dH_2O was then poured onto the membrane and four pieces of Whatman 3M filter paper were placed on top of the membrane. A stack of paper towels were then placed on top of the Whatman filters and a 200g weight was placed on top of construction. DNA was then blotted overnight for 12-16 hours. Following this, the blotting apparatus was disassembled and the membrane was washed for 20 minutes in 2xSSC. The membrane was then air dried and baked at 120°C for 40 minutes to immobilise the DNA.

2.2.7.3 P^{32} labelling of DNA probes

For details of primers used to generate probes used for southern blotting see Table 2.1. All probes were generated by PCR as described in section 1.2.3. The PCR products were then purified using Quiagen PCR purification columns following the manufacturers instructions. 25ng of probe was then radiolabelled with dCTP-^{32}P using the Rediprime II kit (Amersham), following the manufacturers instructions. Labelled probe was then separated from unincorporated nucleotides by passing through ProbeQuant G-50 Micro Columns (Amersham).

2.2.7.4 Hybridisation of Southern blots

Hybridisations were performed in Techne hybridisation bottles rotating in a Techne HB-1 oven using 20ml of Rapid-hyb buffer (Amersham). Pre-hybridisation, hybridisation and washed were carried out at 65°C.

Southern blots were pre-hybridised for 15 minutes in Rapid-hyb buffer. 25ng of radiolabelled probe was denatured at 100°C, snap cooled on ice for 5 minutes and added directly to the hybridisation buffer. Blots were hybridised for 2-3 hours. Following hybridisation, blots were washed for 2x15 minutes in 2xSSC; 0.1% SDS, and 2x15 minutes in 1xSSC; 0.1% SDS. After washed, blots were wrapped in Saran wrap and exposed either to autoradiographic film at -70°C for 4 days or to phosphorimaging plates.

2.3 Tissue Culture

2.3.1 ES cell culture

All cell manipulations were performed in laminar flow sterile hoods using a sterile technique that included wiping the hood down and spraying all items entering the hood with 70% industrial methylated spirits. Cell culture plastic ware was supplied by Iwaki. All cell culture flasks were gelatinised (5 minutes, 0.1% gelatin in PBS) prior to addition of ES cells. E15Tg2a ES cells were used for all ES cell experiments and COS cells (Simian fibroblasts) for transient transfection of *Rhox4* over-expression constructs. Cells were incubated in 7.5% CO_2 at 37°C in a humidified incubator (Scientific Laboratory Supplies Ltd). All solutions were tested for sterility and warmed to 37°C prior to use. ES cells were examined using an inverted microscope (Olympus CK2).

2.3.1.1 Reagents

ES and COS cells were maintained in 1 x glasgow minimum essential medium (GMEM) (GibcoBRL) containing: 10% foetal calf serum; 15% sodium biocarbonate;

0.1% nonessential amino acids; 4mM glutamine; 2mM sodium pyruvate; 0.1mM 2-mercaptoethanol. ES cells were maintained in an undifferentiated state by addition of 100x leukaemia inhibitory factor (LIF).

2.3.1.2 Passage and expansion of cells

Cultures were monitored every day to ensure that cells had not become over-confluent. Cells were normally passaged every 2 days. Culture medium was aspirated and cells were rinsed twice in PBS. Trypsin solution (0.025% trypsin (Gibco); 0.1% chicken serum (Flow Labs); 1.3 mM EDTA in PBS) was added to the cells and incubated at 37°C for 5 minutes until cells had detached. The trypsin was quenched with at least 5ml medium and the cells dispersed to a single cell suspension by gentle pipetting with a narrow bore glass pipette. The cell suspension was centrifuged at 1200rpm for 3 minutes. Following this, the supernatant was aspirated and the cell pellet was resuspended in medium. Cells were then transferred to fresh wells or flasks at dilution ratios of 1:2 or 1:8 depending on rates of growth.

2.3.1.3 Transient transfections

Transient transfections on E14Tg2a ES cell line and COS cell line were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. ES cell transfections were performed in 6-well plates with 1μg of eGFP-C1 over-expression plasmid (Invitrogen) and 3μg of shRNA expression plasmid. COS cell transfections were performed in 100cm cell culture dishes with 24μg of *Rhox4* over-expression plasmid.

2.3.1.4 ES cell differentiation protocols

ES cells were trypsinised and centrifuged to pellet as above. Cells were then resuspended in media (without LIF) and diluted to a concentration of 10^5 cells/ml. Embryoid bodies (EB) were formed by pipetting 10μl drops of the cell suspension onto the inside of a 100cm bacteriological plate lid. The lid was subsequently

inverted onto a base containing 20ml PBS so that the drops are hanging in a humid environment and the EBs were cultured for 48 hours. The EBs were then collected by flooding the lid with media, removing to a conical bottomed 20ml tube for 2 minutes to allow the EBs to settle by gravity and the media was replaced. For a suspension culture (called – gelatin) the EBs were transferred to a new 100cm bacteriological plate in 20ml media without LIF. For adherent culture the EBs were resuspended in 20ml media without LIF and plated onto a 100cm cell culture plate coated with gelatin (5 minutes, 0.1% gelatin in PBS). Both adherent and suspension cultures were cultured for 5 days and the media was changed on every second day either by aspiration of culture medium (adherent cultures) or by allowing the EBs to settle by gravity (suspension cultures). 1 plate of EBs was harvested each day by centrifugation, the cells resuspended in 1ml Trizol and processed as for RNA isolation.

2.4 Flow cytometry

2.4.1 Cell preparation

For separation of transiently transfected cells into GFP positive and negative fractions, cells were removed from monolayer culture by trypsinisation and were resuspended as single cells in FACS wash (10% foetal calf serum in PBS) at 1 x 10^7 cells/ml. Before flow cytometric cell sorting a 1:5000 dilution of Topro3 (Molecular Probes) was added to the cell suspension, to enable exclusion of dead cells from the sorted populations.

Preparation of CD45 positive cells and CD45 negative/MTS20 negative fraction of E15.5 thymic lobes was carried out by N. Blair. For separation of E15.5 thymic lobes, lobes were dissected into PBS. After dissection the PBS was removed and 1ml cell dissociation mixture added (2mg/ml hyaluronidase; 0.7mg/ml collagenase; 0.05mg/ml deoxyribonuclease in PBS) and lobes were incubated at 37°C for 10 minutes. The dissociation mixture was removed after centrifugation at 1200rpm for 3 minutes and PBS/trypsin solution added (0.025% trypsin (Gibco); 0.1% chicken

serum (Flow Labs); 1.3 mM EDTA in PBS), lobes were then incubated at 37°C for a further 10 minutes. To pellet, lobes were centrifuged at 1200rpm for 3 minutes and washed in 2ml FACS wash (10% FCS in PBS). Cells were then incubated with primary antibodies; MTS20 (provided as hybridoma supernatants by R. Boyd, Monash University Medical School, Melbourne, Australia); CD45-FITC (Pharmingen); Ter119-biotin (Pharmingen) (enables red blood cells to be excluded)) for 10 minutes at 4°C, washed with FACS wash. Subsequently they were incubated with secondary antibody (anti-rat IgM-PE (Pharmingen); rabbit anti-rat IgG-FITC(Fcγ) (Jackson Labs); strep-FITC (Pharmingen)) for 10 minutes at 4°C, washed in FACS wash, resuspended in 1.5ml FACS wash and stained for exclusion of dead cells with Topro3 as above.

2.4.2 Cell sorting

Cell sorting was carried out by J. Vrana. Single cell suspensions were prepared and stained as described above. Cells were sorted on a FACS Star machine (Becton Dickinson), gating was applied to exclude dead cells and sorted populations were re-analysed by FACS to assess purity.

2.5 Animals

All animals were housed and treated in accordance with the animal (scientific procedures) act 1986. Animals were housed in a 12:12 light dark cycle with food and water provided *ad libitum.*

Foxn1/LacZ mice used to generate *nude* embryos for *in situ* hybridisation were derived from ES cells that were a generous gift from T. Boehm (MPI Immunobiologie, Frieburg, Germany) (Nehls, Kyewski et al. 1996). The colony was maintained by crossing heterozygous males with C57Bl/6 females.

Gcm2 (Gunther, Chen et al. 2000) and *Hoxa3* (Condie and Capecchi 1993) null embryos were a generous gift from N. Manley (UGA, Athens, USA) and were received fixed and dehydrated for use in *in situ* hybridisation.

2.5.1 Embryo collection

Routinely, matings were between C57BL/6 females and CBA males or between DA rat females and DA rat males. Animals were caged together overnight and females examined for the presence of a vaginal plug the following morning. This was taken as embryonic day 0.5 (E0.5). Pregnant females at the desired stage were sacrificed and the uterus dissected into PBS. Embryos were removed from the uterus and all extraembryonic tissues removed. Embryos were staged by somite counts or by morphological criteria.

2.6 Tissue isolation

Embryos and embryonic thymi were removed by microdissection under a dissecting microscope (Olympus SZ40). Adult tissues were dissected free of connective tissue and washed in PBS.

2.6.1 Endoderm dissection

Isolation of the endodermal gut tube was carried out under an Olympus SZH dissecting microscope with the aid of a pulled glass pipette. E9.5 embryos were removed from the uterus into M2 medium (Sigma) and freed from extraembryonic tissues. Structures caudal to the heart were removed by dissection and the remaining tissue digested in a pancreatin / trypsin enzyme mixture for 15-20 minutes on ice then transferred to M2 medium. Following removal of the head, the ectoderm was peeled off and the heart and neural tube were removed. The remaining mesenchyme was then stripped away to leave a clean gut tube that was trimmed to include on the 3rd pharyngeal pouch. Tissue was stored in Trizol reagent (Invitrogen) for RNA extraction.

Trypsin/Pancreatin Enzyme Mix

0.5g pancreatin (from porcine pancreas, Sigma)

0.1g trypsin (From porcine pancreas, Sigma)

0.1g polyvinylpyrrolidone (PVP) (Sigma)

In 20ml Ca^{2+}-Mg^{2+}-free Tyrode Ringer's Saline

The solution was filtered through a No.1 Whatman filter before filter-sterilising through a 0.45μm syringe filter (Sartorius), and stored at -20°C in 1ml aliquots

Ca^{2+}-Mg^{2+}-free Tyrode Ringer's Saline

8g NaCl

0.3g KCl

0.093g $NaH_2PO_4.5H_2O$

0.025g KH_2PO_4

1g $NaHCO_3$

2g Glucose

Made up to a final volume of 1 litre with sterile water, and pH adjusted to pH7.6 – pH7.7.

2.7 Human embryos

First trimester human embryos were obtained from with ethical approval from We have the following approval for work with human foetal tissue, from both the Lothian University Hospitals NHS Trust and the Lothian Research Ethics Committee: Smith et al "Isolation and propagation of fetal stem cells" LREC/2002/6/15. Consent is obtained from all donors, and the tissue is anonymized before being made available to us. Use and disposal of tissues are strictly regulated. Consenting is done by a fully trained research nurse.

2.8 *In situ hybridisation*

2.8.1 Probes used for *in situ* hybridisation

All antisense riboprobes were DIG-labelled and were used at a concentration of 1μg/ml in hybridisation buffer. Details of *in situ* probes used in this thesis are presented in Table 2.2.

Table 2.2 Details of probes used for *in situ* hybridisation

Probe	Description and Preparation	Source
mRhox4	Full-length sequence cloned from d5 cDNA library. 893bp fragment. Linearised with *Xho1*, transcribed with T3	G. Graham, Beaston Institute for Cancer Research, UK
mGcm2	927bp fragment corresponding to 3' end exon. Linearised with *EcoRI,* transcribed with T7	D. Anderson, Howard Hughes Medical Institute, USA
mFoxn1	500bp exon 3 fragment. Linearised with *HindIII,* transcribed with T7	C. Blackburn, ISCR
mPax1	313bp paired-box fragment. Linearised with *HindIII,* transcribed with T7	P. Gruss, Max Plank Institute, Germany
mHoxa3	Exon 1/2 fragment. Linearised with *XhoI*, transcribed with T3	N. Manley, UGA, Athens, Georgia
hFoxnI	365bp fragment. Linearised with *PmeI,* transcribed with T7	Cloned from 61 week thymic RNA
hHoxa3	379bp fragment. Linearised with *NotI,* transcribed with T3	Cloned from 61 week thymic RNA
hPaxI	512bp fragment. Linearised with *PmeI*, transcribed with T7	Cloned from 61 week thymic RNA

2.8.1.1 Probe synthesis reaction

Plasmids containing the coding sequence were linearised and used as templates for riboprobes. Restriction digests were performed using standard protocols, as recommended by the suppliers (NEB and Roche). Digestion products were routinely analysed by agarose gel electrophoresis with gels cast and run in 1 x TAE buffer containing 0.5μg/ml ethidium bromide. Probe synthesis incorporated DIG-UTP.

1μg purified linear template
2μl NTP labelling mix (Roche)
2μl 10x transcription buffer (Roche)
1μl RNase inhibitor (Roche)
2μl RNA polymerase (Roche)

Made up to a final volume of 20μl in RNase-free water.

Reactions were incubated for 2 hours at 37°C, then stopped by adding 0.2M EDTA and precipitated with 4M lithium chloride. Riboprobes were stored at -20°C at 0.1mg/ml.

Hybridisation buffer
1% Boehringer block
50% formamide
5 x SSC
1mg/ml yeast RNA
0.1mg/ml heparin
0.1% Tween-20
0.1% CHAPS
5mM EDTA
In RNase-free water

The solution was dissolved by heating to 50°C and stored at -20°C.

2.8.2 Wholemount *in situ* hybridisation

Embryos were removed from the uterus, freed of all extraembryonic tissues and fixed overnight in 4% PFA in PBS at 4°C. The following day embryos were dehydrated through a methanol series and stored in 100% methanol at -20°C until required.

2.8.2.1 Pre-hybridisation

Prior to rehydration, head cavities were pierced in E9.5 and E10.5 embryos to avoid trapping of probe and staining substrate. Embryos were rehydrated through a methanol series, washed in PBST (0.1% Tween-20 in PBS), and bleached for 1 hour in 6% hydrogen peroxide (Sigma). After washing, tissue was permeabilised by incubation with 10μg/ml Proteinase K (Sigma), as appropriate: 5 minutes for E8.5, 15 minutes for E9.5, 30 minutes for E10.5. Proteinase K was inactivated by two 5-minute incubations in 2mg/ml glycine and tissues were washed and refixed in 4% PFA/0.2% glutaraldehyde. Embryos were then incubated for 1 hour in hybridisation buffer at 70°C to block non-specific binding.

2.8.2.2 Hybridisation

Embryos were hybridised at 70°C in 1μg/ml DIG-labelled probe in hybridisation buffer.

2.8.2.3 Post-hybridisation

After hybridisation, embryos were washed for: three times 30 minutes in Solution 1 (50% formamide; 5xSSC (pH4.5); 1% SDS) at 70°C, 10 minutes in 50:50 Solution 1:2 at 70°C, three times 5 minutes in Solution 2 (0.5M NaCl; 10mM Tris-HCl (pH7.5); 0.1% Tween-20) at room temperature, two times 30 minutes in 100μg/ml RNaseA (Sigma) in Solution 2 at 37°C, three times 30 minutes in Solution 3 (50% formamide; 2xSSC) at 65°C. Embryos were then washed for three times 10 minutes in TBST and blocked in 10% sheep serum for 60-90 minutes before incubating overnight in 1:2000 anti-DIG-AP antibody at 4°C.

The following morning embryos were washed for three times 5 minutes then five times 1 hour in TBST, and stored in TBST at 4°C overnight. Prior to the colour reaction, embryos were washed 3x20 minutes in freshly prepared NTMT (100mM NaCl; 100mM Tris-HCl (pH9.5); 50mM $MgCl_2$; 0.1% Tween-20; 2mM Levamisole). 1ml BM Purple (Roche) was added to each sample, and samples were incubated in the dark at room temperature until the desired colour developed. If required, staining was continued overnight at 4°C. Embryos were then washed in PBS, post-fixed overnight in 4% PFA at 4°C, and stored in PBS; 0.05% EDTA at 4°C until required.

2.8.2.4 Paraffin embedding and sectioning

R. Wilkie carried out paraffin embedding and sectioning of human embryos processed by *in situ* hybridisation. Embryos for paraffin embedding were treated as follows:

70% ethanol	15 minutes
90% ethanol	15 minutes
100% ethanol	15 minutes
100% ethanol	15 minutes
xylene	15 minutes
xylene	15 minutes
wax (60°C)	30 minutes
wax (60°C)	30 minutes
wax (60°C)	30 minutes

Samples were embedded in wax and cooled to 4°C. Sections were cut at 7μm. Wax was removed by two 4 minutes incubations in zylene followed by mounting in DPX.

For photography, images were captured using 100ASA print film (Fujifilm) and an SZH10 stereomicroscope (Olympus) with C-35AD-4 camera (Olympus).

2.9 Bioinformatics

2.9.1 Sequence alignments

Sequence alignments were carried out using the default settings of the Clustal W algorithm, MegAlign version 6.0 (DNASTAR). AVID global alignments were created using the AVID program of the mVISTA package using the default settings (http://genome.lbl.gov/vista/index.shtml) (Bray, Dubchak et al. 2003).

2.9.2 *Ka/Ks* calculations

FASTA sequence alignment files were generated using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Calculation of the number of synonymous and non-synonymous substitutions between *Rhox2, 3* and *4* alleles were then carried out using dnaSP version 3 with the default settings (http://www.ub.es/dnasp/) (Rozas and Rozas 1999).

Chapter 3 Expression of *Rhox4* in Thymus Organogenesis[1]

3.1 Introduction

The thymus plays a central role in the vertebrate immune system. However, despite its importance the genetic control of early thymic development remains poorly understood. The thymus arises in a common primordium with the parathyroid gland, which develops from 3rd pharyngeal pouch endoderm between day 9.5 and 11.5 of murine embryogenesis (E9.5 and E11.5 respectively) (Schreier and Hamilton 1952; Hammond 1954). It is likely that the transcription factor *Gcm2* is responsible for establishment of the parathyroid region of the 3rd pharyngeal pouch, as this gene is expressed in the dorsal region of the 3rd pharyngeal pouch from E9.5 (Gordon, Bennett et al. 2001) and *Gcm2* null mice have no parathyroid glands (Gunther, Chen et al. 2000). The first gene known to solely affect thymus organogenesis is the transcription factor *Foxn1,* a member of the forkhead family of proteins. Mutations in this gene result in athymia in mice and humans (Nehls, Pfeifer et al. 1994; Frank, Pignata et al. 1999). *Foxn1* is expressed in the thymic primordium from E11.25 (Gordon, Bennett et al. 2001) and is required cell autonomously for TEC development (Blackburn, Augustine et al. 1996). However, three lines of evidence suggest that *Foxn1* is not required for specification of the thymus; 1) *nude* mice contain a thymic rudiment, and thymus development in *nu/nu* mice is affected only from around E12, subsequent to formation of the thymic primordium (Flanagan 1966), 2) grafting of E8.5 – E9.0 2nd and 3rd pouch endoderm under the kidney

---

[1] This work represents a continuation of initial observations made in this lab by Julie Gordon. The expression of *Rhox4* (then *Ehox*) in the developing thymus was published after work described here was completed, by
Jackson, M., J. W. Baird, et al. (2003). "Expression of a novel homeobox gene Ehox in trophoblast stem cells and pharyngeal pouch endoderm." Dev Dyn **228**(4): 740-4.

capsule of nude mice gives rise to a functional thymus (Gordon, Wilson et al. 2004), and 3) in the pharyngeal region, IL-7, which is required for early T cell development, is expressed specifically in the 3rd pharyngeal pouch from E11.5, in a *Foxn1* independent manner (Zamisch, Moore-Scott et al. 2005).

Phenotypic analysis of single and compound mouse mutants has suggested a transcription factor hierarchy required for thymus organogenesis, this includes members of the *Hox, Pax* and *Eya* families. However, whilst *Hoxa3, Pax1, Pax9* and *Eya1* obviously play important roles in thymus development, they also regulate development of the parathyroid gland: *Hoxa3*$^{-/-}$, *Pax1*$^{-/-}$, *Pax9*$^{-/-}$ and *Eya1*$^{-/-}$ mice all have defects in parathyroid gland formation (Dietrich and Gruss 1995; Manley and Capecchi 1995; Hetzer-Egger, Schorpp et al. 2002; Xu, Zheng et al. 2002). *Gcm2* expression is restricted to the parathyroid domain of the 3rd pharyngeal pouch two days before initiation of high level *Foxn1* expression in the thymic domain (Gordon, Bennett et al. 2001). Thus, either thymus is the default pathway for the 3rd pharyngeal pouch endoderm or there are other, unidentified genes involved in specification of the thymus and induction of *Foxn1* expression.

Rhox4 (formerly *Ehox*) is a member of the *Rhox* family of homeodomain-containing transcription factors which are primarily expressed in reproductive and extra-embryonic tissues (Maclean, Chen et al. 2005). Uniquely amongst *Rhox* family members, *Rhox4* is expressed in the thymus (Jackson, Baird et al. 2002; Maclean, Chen et al. 2005), and initial observations made in this lab indicated that *Rhox4* is expressed during thymus organogenesis (J. Gordon). In this chapter, I describe *Rhox4* expression during thymus development. Furthermore, I analyse *Rhox4* expression in *Hoxa3, Gcm2* and *Foxn1* mutant embryos to determine whether *Rhox4* expression is regulated by transcription factor pathways with established roles in thymus organogenesis.

3.2 Results

3.2.1 Detailed analysis of the spatial and temporal expression profile of *Rhox4* during thymus ontogeny.

To determine whether *Rhox4* is expressed during specification of the thymus and subsequent organogenesis, *in situ* hybridisation was carried out on staged mouse embryos using a probe corresponding to full length *Rhox4* (Jackson, Baird et al. 2002). *Rhox4* expression was first detected at E8.5 in the ventral foregut endoderm (Figure 3.1a,b). Between E8.5 and E9.5, the time frame in which the first three pharyngeal pouches form, *Rhox4* is expressed transiently in the ventral regions of all three pouches and at a low level in the region of the foregut where the 4th pouch will form (Figure 3.2a). The *Rhox4* expression domain is then progressively restricted such that by E9.5 expression is limited to the ventral region of the 3rd pharyngeal pouch, the region destined to become thymus (Figure 3.2b). *Rhox4* was expressed at barely detectable levels in the 3rd pharyngeal pouch at E10.5 (Figure 3.3a,b), and by E11 expression could not be detected by *in situ* hybridisation (Figure 3.4a). Strikingly, *Foxn1* is expressed in the former *Rhox4* domain from E11.25 (Figure 3.4b) (Gordon, Bennett et al. 2001), suggesting a regulatory relationship between the two genes. *Rhox4* expression was not detected in any other regions of the embryo at any of these time points.

To further define the expression pattern, *Rhox4* expression was compared with other genes known to be present in the region; *Hoxa3, Pax1* and *Gcm2*. Initiation of *Rhox4* and *Pax1* expression appears to occur simultaneously at E8.5 in the same region of the ventral foregut endoderm (Figure 3.1c). Subsequently, *Pax1* is expressed throughout the four pharyngeal pouches, although staining is less intense at the dorsal/anterior tip at E9.5 (Figure 3.2d) (Muller, Ebensperger et al. 1996; Wallin, Eibel et al. 1996). The homeobox gene, *Hoxa3* is initially expressed in the dorsal mesoderm, at E8.5 (Figure 3.1c). Uniquely amongst Hox3 paralogues, from E9.5, expression is found throughout the 3rd pharyngeal pouch endoderm in addition to the

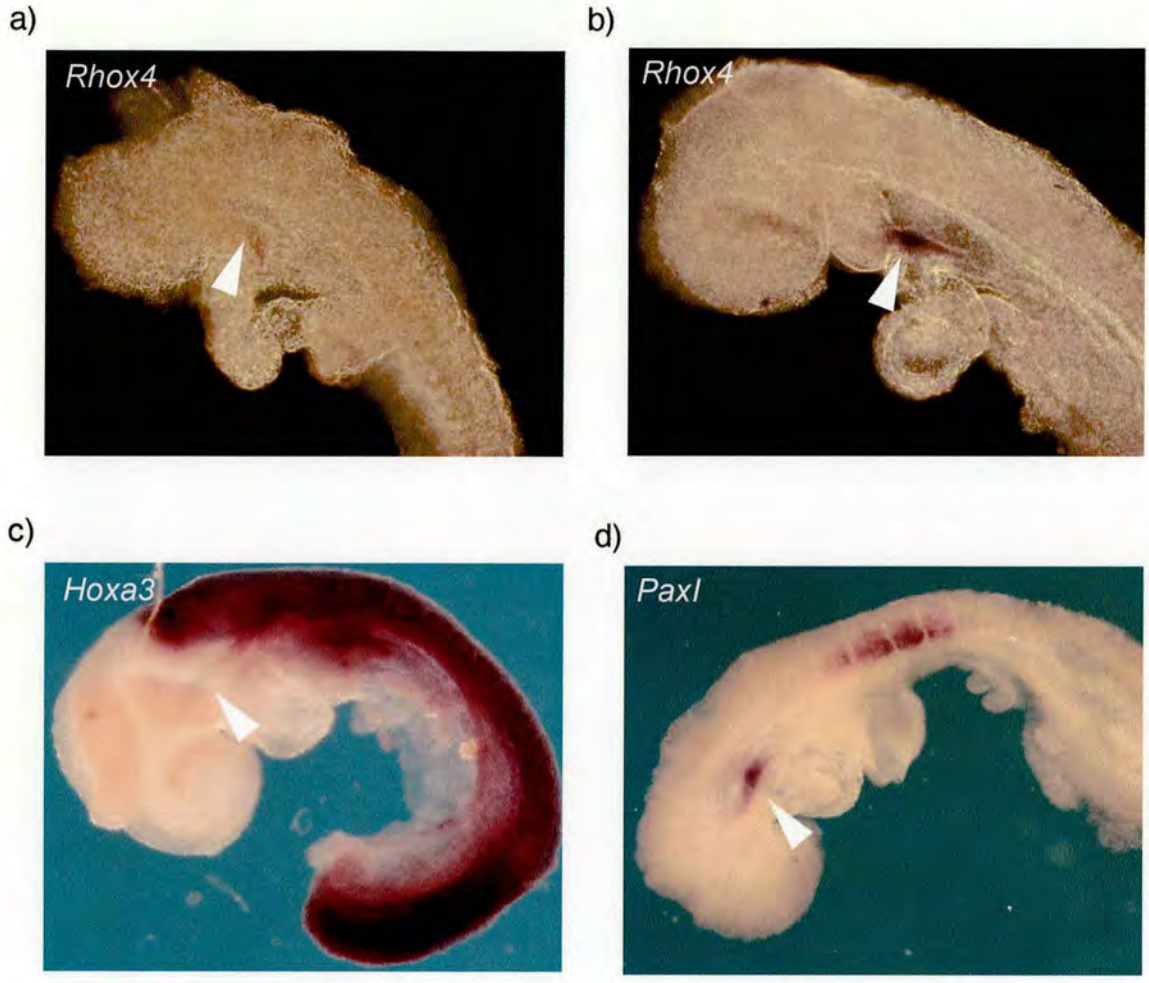


Figure 3.1 *Rhox4* is expressed in the ventral foregut endoderm from E8.5 in the same domain as *Pax1* but before onset of *Hoxa3* expression. Images show whole embryos processed by *in situ* hybridisation using probes for *Rhox4* (a,b), *Hoxa3* (c), *Pax1* (d). a) No *Rhox4* expression was detected at E8.0; b) by E8.5, *Rhox4* marked a domain in the ventral floor of the foregut endoderm; c) *Hoxa3* is expressed in the neural crest derived mesenchyme but is not detected in the foregut endoderm at E8.5; d) *Pax1* is expressed in the ventral floor of the foregut endoderm in an E8.5 embryo, some expression is also seen in the developing somites. Arrows indicate the position of the ventral foregut endoderm. Orientation: head is left and dorsal is up.

Figure 3.2 *Rhox4* expression is restricted to the ventral region of the 3rd pharyngeal pouch by E9.5. Images show whole embryos processed by *in situ* hybridisation using probes for *Rhox4* (a,b), *Hoxa3* (c), *Pax1* (d) and *Gcm2* (e). a, b) At E9.0, *Rhox4* is expressed in the ventral region of pharyngeal pouches 1, 2 and 3, and by E9.5 is restricted to the ventral domain of the 3rd pharyngeal pouch; c) At, E9.5, *Hoxa3* is detected in the 3rd pharyngeal pouch and more caudally; d) *Pax1* expression at E9.5 is seen in the pharyngeal pouches 1 to 3 and e) *Gcm2* is expression is initiated at E9.5 in the dorsal region of the 3rd pharyngeal pouch. p1, 2, 3: pharyngeal pouches 1 to 3 respectively. Orientation: head is left and dorsal is up.

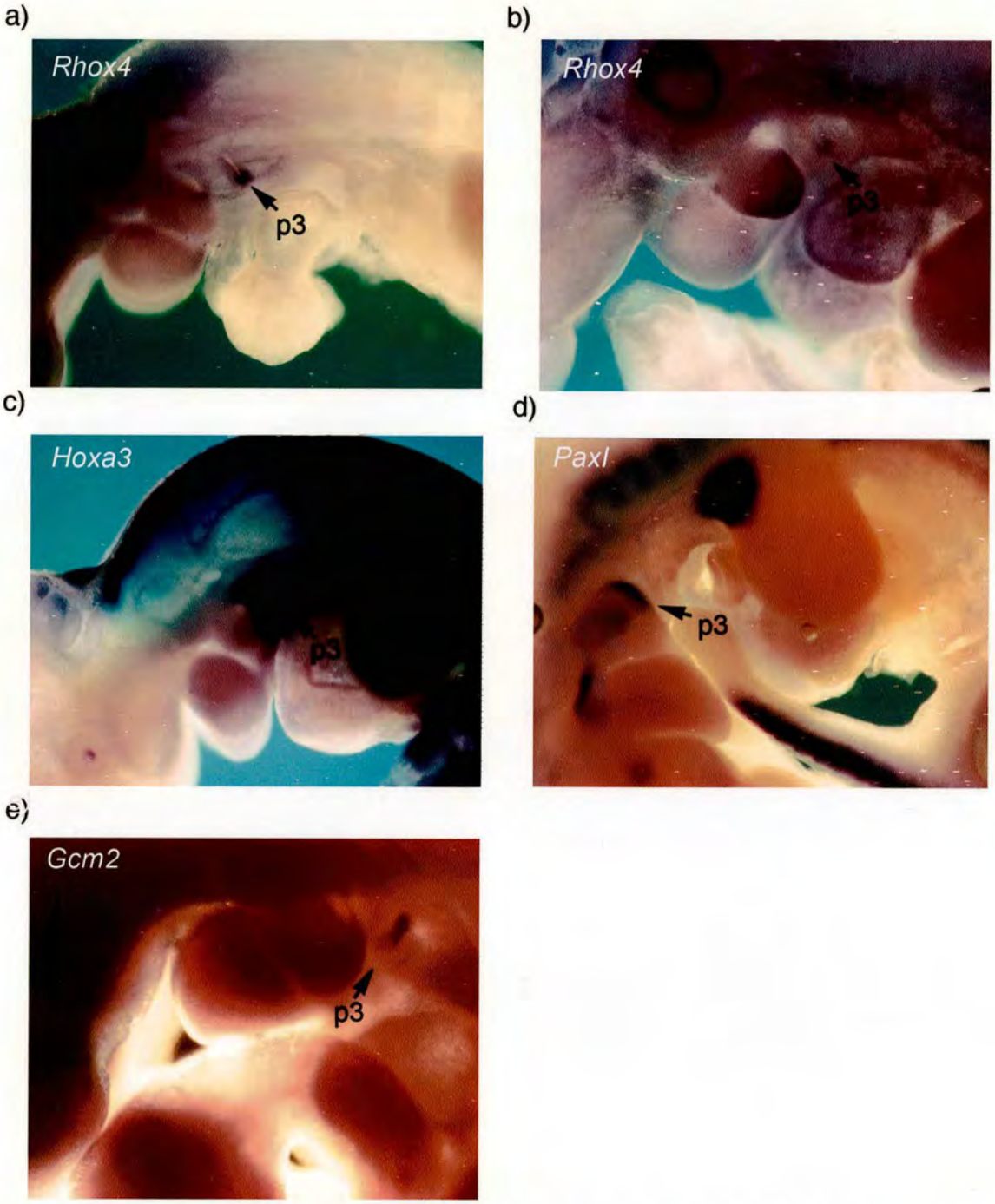


Figure 3.3 *Rhox4* expression is being down-regulated at E10.5 in the ventral region of the 3rd pharyngeal pouch. Images show whole embryos processed by *in situ* hybridisation using probes for *Rhox4* (a,b), *Hoxa3* (c), *Pax1* (d) and *Gcm2* (e). a, b) *Rhox4* expression is present at E10.0 (a) and E10.5 (b) in the ventral portion of the 3rd pharyngeal pouch however, the level of expression is lower at E10.5 (b); *Hoxa3* (c) and *Pax1* (d) are still expressed in the thymic primordium at E11.0; e) *Gcm2* expression in the dorsal region of the 3rd pharyngeal pouch at E10.5. p3 indicates the position of the 3rd pharyngeal pouch. Orientation: head is left and dorsal is up.

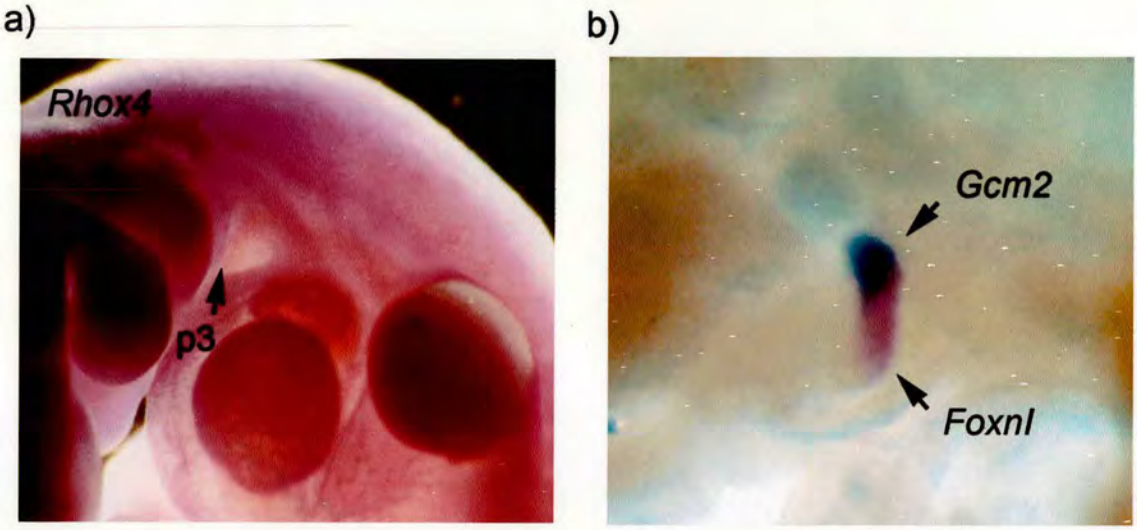


Figure 3.4 *Rhox4* expression can not be detected in the thymic primordium at E11.5. Images show whole embryos processed by *in situ* hybridisation using probes for a) *Rhox4* in a E11.5 embryo, showing no expression in the ventral region of the 3rd pharyngeal pouch, the position of the 3rd pharyngeal pouch is indicated (p3); b) *Foxn1* (red) and *Gcm2* (purple), taken from (Gordon et al. 2001), showing expression in complementary domains of the 3rd pharyngeal pouch at E11.5. Orientation: head is left and dorsal is up.

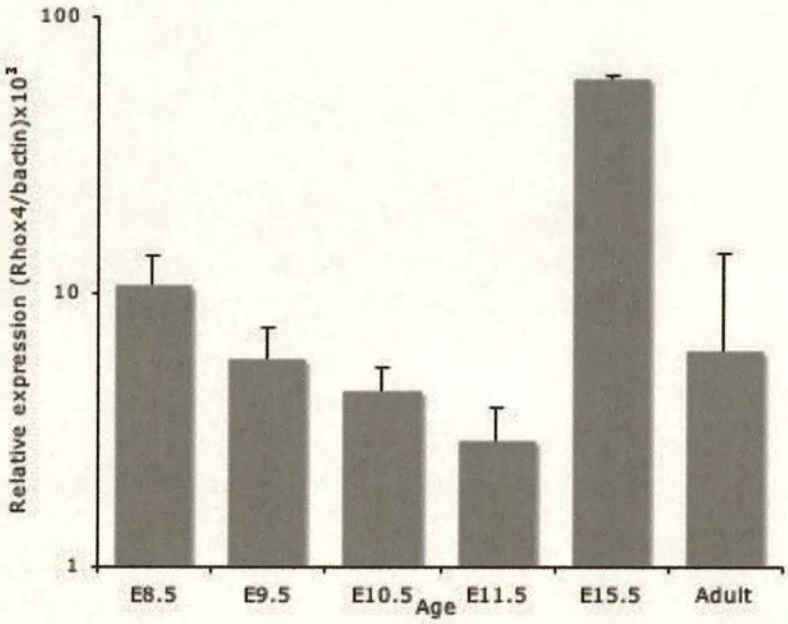
surrounding mesenchyme (Figure 3.2c) (Manley and Capecchi 1995; 1998). The anterior and dorsal region of the 3rd pharyngeal pouch gives rise to the parathyroid gland. Expression of the parathyroid-specific gene, *Gcm2* marks this region from E9.5, the point at which the 3rd pharyngeal pouch is formed (Figure 3.2e, 3.3e) (Gordon, Bennett et al. 2001). Expression overlaps with *Rhox4* in a small region of the anterior pouch at E9.5. However, by E10 *Rhox4* and *Gcm2* are expressed in complementary regions of the 3rd pharyngeal pouch. *Gcm2* expression is sustained in the dorsal pouch (Gordon, Bennett et al. 2001) after *Rhox4* has been strongly down-regulated (Figure 3.3e, 3.4b).

A previous study has indicated that *Rhox4* is expressed in the neonatal and adult thymus, in addition to earlier expression in the foregut and 3rd pharyngeal pouch endoderm (Jackson, Baird et al. 2002; Maclean, Chen et al. 2005). As *Rhox4* expression could not be detected by *in situ* hybridisation after E10.5, we carried out quantitative RT-PCR using cDNA from stages of thymus development from E8.5 through E15.5 and adult (Table 3.1, Figure 3.5a). In these analyses, RNA was obtained from micro dissected whole thymi; however, at E8.5, E9.5, E10.5 and E11.5 dissections all contained tissue surrounding the endoderm or thymus primordia and, as *Rhox4,* is highly expressed only in the endoderm (Table 3.1, Figure 3.5b) the level of expression in these samples appears artificially low. *Rhox4* expression was detectable throughout ontogeny and in the adult thymus. Surprisingly a high level of expression was detected at E15.5, as compared to both E11.5 and adult. Since lymphoid precursors infiltrate the thymic primordium from around E11 (Moore and Owen 1967; Le Douarin and Jotereau 1975), it was possible that the increase in *Rhox4* expression in the thymus after E11 reflected expression by lymphocytes. RT-PCR was therefore carried out on RNA isolated from the lymphocyte (CD45 positive) fraction of E15.5 lobes after purification of this fraction by flow cytometry (carried out by N. Blair). In addition RT-PCR was carried out on both intact E15.5 lobes, and the population containing differentiated TEC and mesenchymal cells, which were identified as the CD45 negative/MTS20 negative fraction. MTS20 is a monoclonal antibody which marks thymic epithelial progenitor cells at E15.5 (Bennett, Farley et al. 2002; Gill, Malin et al. 2002), the CD45 negative/MTS20

Table 3.1 *Rhox4* expression throughout thymus organogenesis as determined by real-time RT-PCR. In these analyses, RNA was obtained from micro dissected whole thymi; however, at E8.5, E9.5, E10.5 and E11.5 dissections all contained tissue surrounding the endoderm or thymus primordia and, as *Rhox4,* is highly expressed only in the endoderm (see E9.5 endoderm), the level of expression in these samples appears artificially low. For E9.5 endoderm, n =2, all other stages, n = 4.

Tissue Age	Rhox4	β actin	Ratio of expression (Rhox4/βactin)x10^3	Standard deviation of ratio
E8.5	40,284	3,850,625	10.6	3.0
E9.5	14,778	2,489,250	5.7	1.7
E9.5 endoderm	61, 885	914,500	67.7	
E10.5	10,247	2,446,250	4.4	1.0
E11.5	2,521	787,600	2.9	1.7
E15.5	55,433	929,025	59.6	7.7
Adult	32,076	5,479,250	6.0	1.6

a)



b)

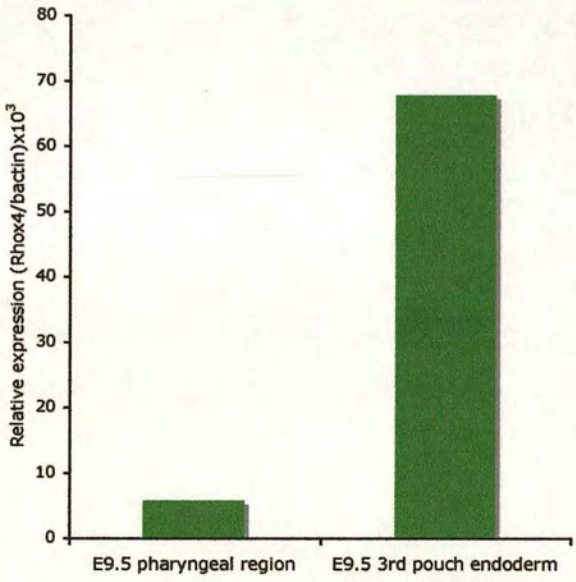


Figure 3.5 *Rhox4* is expressed in the 3rd pharyngeal pouch endoderm and thymic rudiment throughout ontogeny and in the adult thymus. Real-time RT-PCR for *Rhox4* expression during (a) thymus development from E8.5, before formation of the 3rd pharyngeal pouch, to adult (n=4) and (b) expression at E9.5 is highly enriched in the 3rd pharyngeal pouch as compared to the E9.5 pharyngeal region, containing the 2nd and 3rd pharyngeal pouch endoderm and surrounding tissue (n=2).

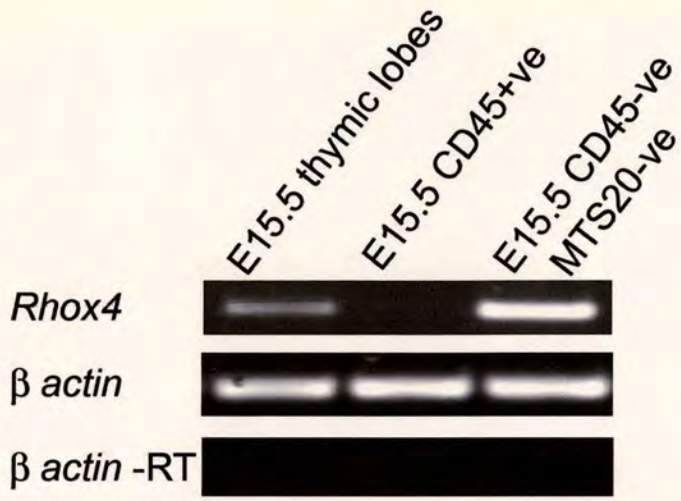


Figure 3.6 *Rhox4* is expressed in the non-lymphocyte compartment of the thymic rudiment at E15.5. RT-PCR (35 cycles) for *Rhox4* and β actin on cDNA synthesised from; intact E15.5 thymic lobes, lymphocytes (CD45 positive) and epithelial/mesenchymal cells (CD45 negative, MTS20 negative). The CD45 negative; MTS20 negative population contains differentiated TEC and mesenchymal cells, but not thymic epithelial progenitor cells, which are MTS20 positive. Cell populations were isolated by staining whole E15.5 thymi followed by flow cytometric cell sorting, and was carried out by N. Blair.

negative population therefore excludes the progenitor cell population. *Rhox4* expression was detected in whole E15.5 lobes and the population containing differentiated TEC and mesenchymal cells. However no expression found in the lymphocyte population (Figure 3.6). *Rhox4* expression at E15.5 is therefore restricted to the non-lymphocyte compartment of the thymic rudiment where it is expressed at a level comparable with the E9.5 endoderm, and may be detectible by *in situ* hybridisation at this stage.

3.2.2 *Rhox4* Expression is Unchanged in Mice with Primary Defects in Thymus Organogenesis

The *Rhox4* expression pattern described above indicates this gene as the first specific marker of the prospective thymic domain of the 3rd pharyngeal pouch, suggesting that *Rhox4* may play a role in specification of this tissue. To determine whether *Rhox4* might function within established transcription factor networks, I therefore investigated whether *Rhox4* expression was altered in mice carrying mutations in genes with key roles in thymic development.

In situ hybridisation using a *Rhox4* specific probe showed no change in *Rhox4* expression in *Hoxa3*$^{-/-}$ embryos compared to wild-type and heterozygous littermates at E9.5 or E10.5 (Figure 3.7). Although, changes occurring at E10.5 would be hard to detect as *Rhox4* expression is being down-regulated at this time, these data suggest that *Rhox4* lies in a different genetic pathway from *Hoxa3* and *Pax1*.

To determine whether in the absence of *Gcm2* the *Rhox4* expression domain expanded to include the dorsal 3rd pharyngeal pouch, as would be expected if *Rhox4* played a role in patterning the pouch into thymus and parathyroid domains, *Rhox4* expression was examined in E9.5 *Gcm2*$^{-/-}$ embryos. No difference in *Rhox4* expression was observed in the *Gcm2* null embryos (Figure 3.8a,b), indicating that *Gcm2* is not required to limit the *Rhox4* expression to the ventral region of the 3rd pharyngeal pouch at E9.5.

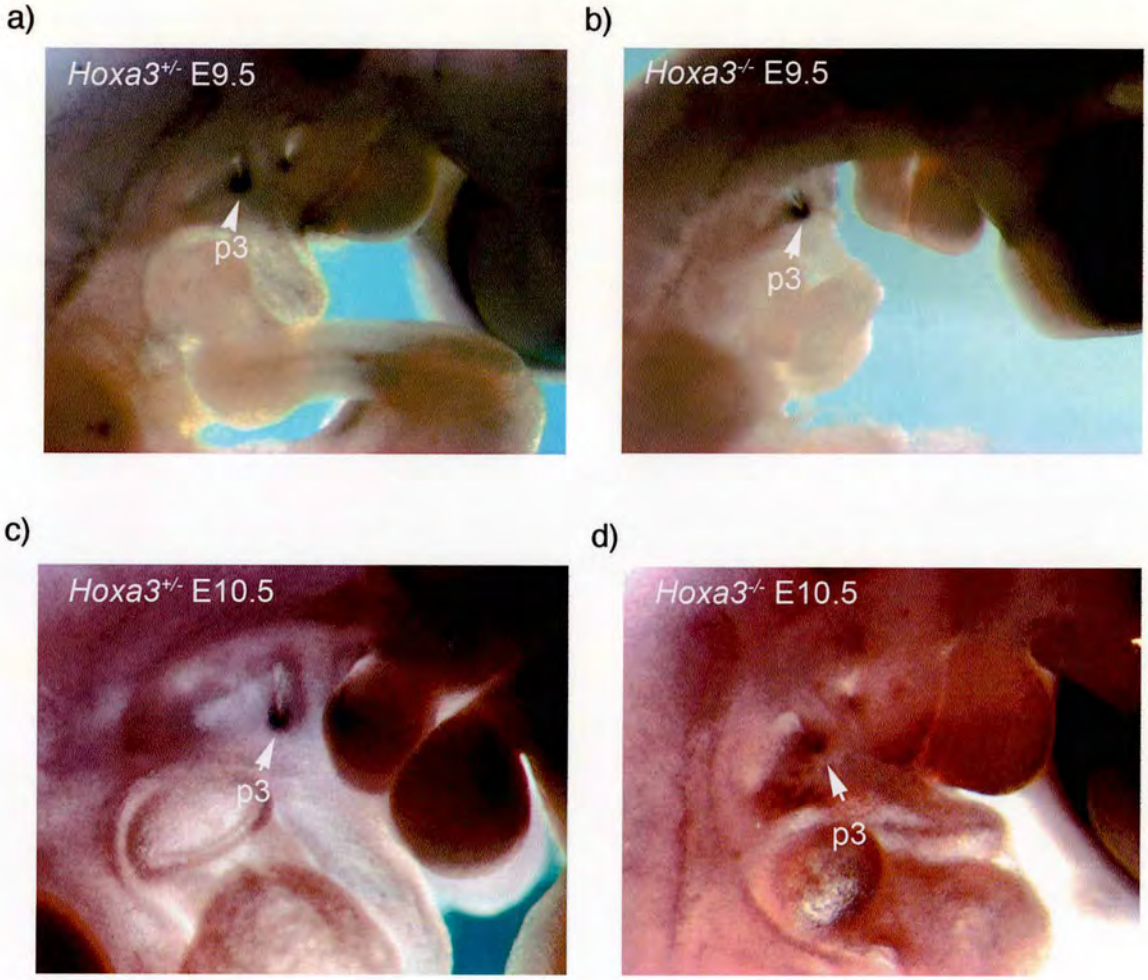


Figure 3.7 *Rhox4* expression is unaffected in *Hoxa3* null embryos. Images show whole embryos processed by *in situ* hybridisation for *Rhox4* in a) E9.5 *Hoxa3*$^{+/-}$; b) E9.5 *Hoxa3*$^{-/-}$; c) E10.5 *Hoxa3*$^{+/-}$; d) E10.5 *Hoxa3*$^{-/-}$. p3 indicates the position of the ventral 3rd pharyngeal pouch, where *Rhox4* is expressed at both E9.5 and E10.5. Orientation: head is right and dorsal is up.

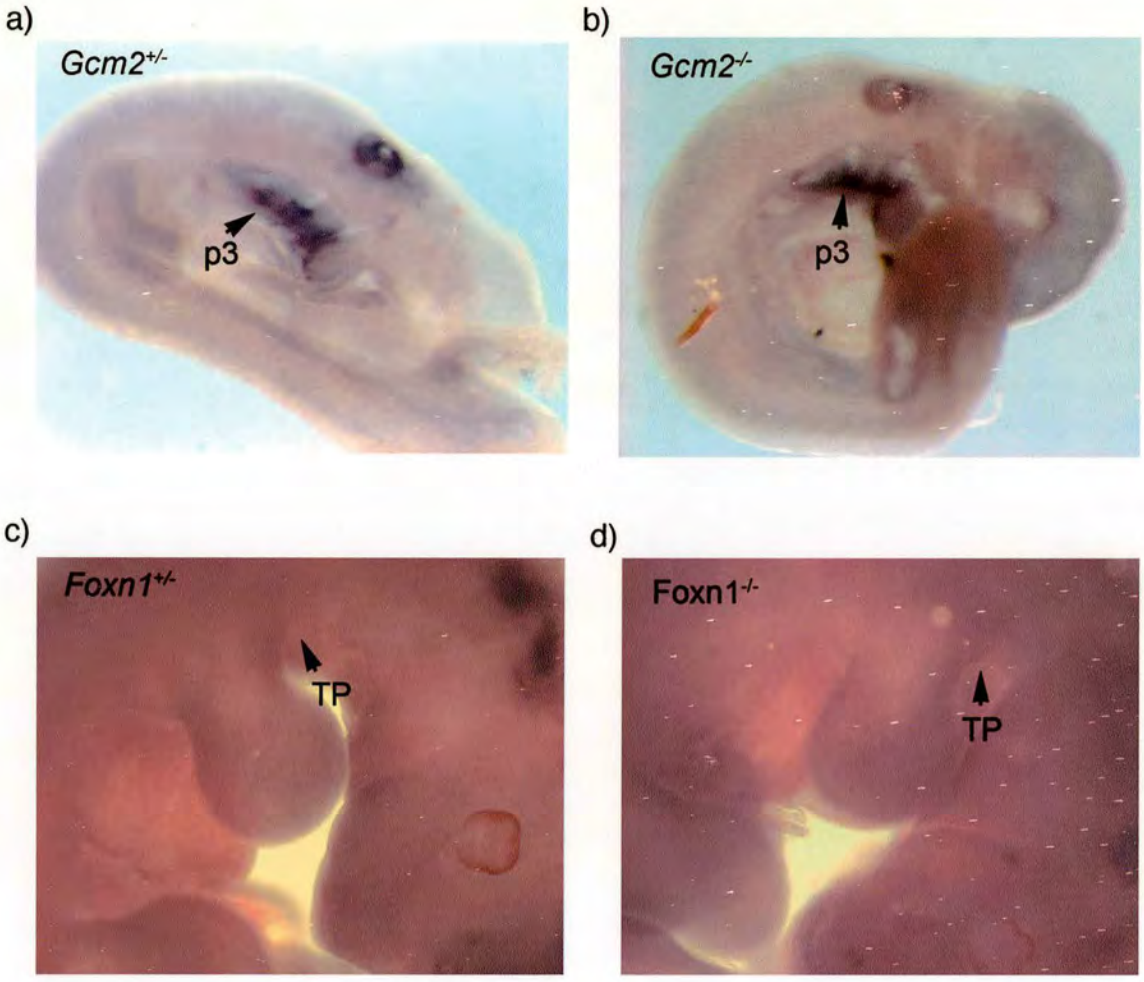


Figure 3.8 *Rhox4* expression is unaffected in *Gcm2* and *Foxn1* null embryos. Images show whole embryos processes by *in situ* hybridisation for *Rhox4* expression in a) E9.5 *Gcm2*$^{+/-}$ embryo; b) E9.5 *Gcm2*$^{-/-}$ embryo; c) E11.5 *Foxn1*$^{+/-}$ embryo; d) E11.5 *Foxn1*$^{-/-}$ embryo. Arrow indicates position of the ventral region of the 3rd pharyngeal pouch (p3) or the thymic primordium (TP). Orientation: head is right and dorsal is up.

The down-regulation of *Rhox4* at the time of high-level *Foxn1* expression suggests a possible regulating relationship for these genes. *In situ* hybridisation for *Rhox4* was therefore carried out on E11.5 *Foxn1*-null embryos to determine whether *Rhox4* expression is prolonged in the absence of *Foxn1*. No *Rhox4* expression was detected in these embryos or littermate controls, indicating that *Foxn1* is not required for down-regulation of *Rhox4* at E11 in the ventral region of the 3rd pharyngeal pouch (Figure 3.8c,d).

3.3 Discussion

In this chapter, I have described detailed analysis of the expression of *Rhox4* during embryonic development. This revealed that *Rhox4* is specifically expressed at E9.5 in the region of the 3rd pharyngeal pouch that is going to form the thymus. Since no other genes have been identified that mark this region before onset of *Foxn1* at E11.25, these data suggest that *Rhox4* may play a role in specification of the thymic domain. Furthermore, *Rhox4* expression was strongly down regulated at the time of onset of *Foxn1* expression suggesting that there may be genetic interaction between these two genes in the 3rd pharyngeal pouch endoderm. Analysis of *Rhox4* expression in embryos null for *Hoxa3, Gcm2* and *Foxn1* however failed to place *Rhox4* in any transcription factor networks with established roles in thymus development.

The progressive restriction of *Rhox4* expression from a broad region of the ventral floor of the foregut endoderm at E8.5 to the ventral portion of the 3rd pharyngeal pouch at E9.5, before onset of *Foxn1,* is striking and strongly suggests that *Rhox4* may play a role in specifying this tissue. Alternatively it may play a role in pharyngeal pouch development or neither. *Rhox4* expression is maintained at a low level throughout ontogeny and in the adult thymus. Thus, *Rhox4* may have additional roles in both later stages of thymus organogenesis and in the postnatal thymus. Surprisingly, the real-time RT-PCR analysis showed that *Rhox4* was highly expressed in the E15.5 embryonic thymus, as compared to E11.5 and adult thymic lobes, and further investigation showed that expression was restricted to the non-lymphoid compartment of the primordia. As the non-lymphoid (epithelial/mesenchymal) component is only 20% of the cellularity at this stage, due to colonisation by lymphoid precursors and T cell proliferation (J.M. Sheridan, unpublished), the actual level of *Rhox4* expression is at least five times that recorded. Further analysis will determine whether expression of *Rhox4* at E15.5 is limited to the epithelial compartment. Distinct cortical and medullary regions can be detected in the thymus from E15.5 (Klug, Carter et al. 1998; Klug, Carter et al. 2002), thus, *Rhox4* may have a role in thymus development during end stage epithelial differentiation.

Thymus organogenesis involves formation of a common thymus/parathyroid primordia, specification of the thymic portion of the 3rd pharyngeal pouch and outgrowth and separation of the thymic and parathyroid regions, followed by colonisation of the thymic rudiment by T cell progenitors and proliferation and maturation of thymic epithelial cells (Blackburn and Manley 2004). Many genes and gene products have been identified that affect one or more stages of thymus organogenesis, although we are currently far from understanding the complete picture. *Rhox4* expression was unchanged in *Hoxa3, Gcm2* and *Foxn1* null embryos. *Hoxa3* is required for the proliferation and outgrowth of the 3rd pharyngeal pouch and null embryos lack both the thymus and parathyroid gland (Manley and Capecchi 1995). Specific loss of *Pax1* expression is also seen at E10.5 in the 3rd pouch endoderm of *Hoxa3*$^{-/-}$ mice. The data presented here show that there is no dramatic change in *Rhox4* expression in *Hoxa3*$^{-/-}$ embryos at E10.5. However, we cannot rule out the possibility that there is a subtle change, since *Rhox4* expression is beginning to be lost in the ventral portion of the 3rd pouch endoderm at this time, making small changes in expression hard to detect by *in situ* hybridisation.

Gcm2 and *Rhox4* are expressed in complementary domains in the 3rd pharyngeal pouch from E9.5, until *Rhox4* expression is lost at E11.5. *Gcm2*$^{-/-}$ mice have no parathyroid glands (Gunther, Chen et al. 2000). As *Rhox4* expression is still restricted to the ventral portion of the pouch in *Gcm2* null embryos at E9.5, either *Rhox4* is responsible for patterning the thymic portion of the 3rd pharyngeal pouch or another factor limits *Rhox4* expression to the ventral portion of the pouch. *Rhox4* expression is down-regulated in 3rd pharyngeal pouch endoderm by E11.5, and *Foxn1* is expressed in its place (Gordon, Bennett et al. 2001). In the absence of *Foxn1*, *Rhox4* expression was not maintained, predicting that *Foxn1* is not responsible for its down-regulation. However the converse scenario is possible. We have therefore been unable to place *Rhox4* in a transcription factor network with an established role in thymus organogenesis. The role *Rhox4* plays in the developing pharyngeal pouches and in thymus organogenesis therefore remains to be determined.

Chapter 4 Organisation of the *Rhox* α locus

4.1 Introduction

The recently described *Rhox* locus comprises three clusters of genes in the A2 region of the mouse X chromosome (Maclean, Chen et al. 2005). The α cluster contains *Rhox1* through to *Rhox4*, the β cluster *Rhox5* to *9* and the γ cluster, *Rhox10, 11* and *12*. Family members are defined by their divergent paired-like homeodomain, which is encoded by three exons. *Rhox* family members are primarily expressed in reproductive and extra embryonic tissues. However, three members are each expressed in one additional tissue, including *Rhox4* - which is expressed in the developing and adult thymus (Chapter 3) (Jackson, Baird et al. 2002; Jackson, Baird et al. 2003; Maclean, Chen et al. 2005).

As described in Chapter 3, *Rhox4* is expressed in the developing pharyngeal endoderm and 3rd pharyngeal pouch in a pattern highly suggestive of a role in early thymus development. An original aim of this thesis was therefore to address *Rhox4* function via generation of conventional and conditional *Rhox4* null copies by gene targeting. As a prerequisite for construct design and building, detailed examination of the *Rhox4* locus structure was carried out. In this chapter, I therefore describe the physical characterisation of the *Rhox* α locus.

4.2 Results

4.2.1 A Seven Copy *Rhox4* Array Exists on Mouse Chromosome X

Rhox4 is present in the A2 region of the mouse X chromosome (Jackson, Baird et al. 2002; Maclean, Chen et al. 2005). The 893bp *Rhox4* cDNA is encoded by four exons, which span 4.7kb of genomic DNA. Although a previous publication reported only one copy of *Rhox4* (Maclean, Chen et al. 2005), alignment of the published *Rhox4* cDNA sequence (NM_021300) with the mouse genome sequence, NCBI build 33, gave multiple, high scoring hits. Further analysis of this sequence suggested the presence of seven complete copies of *Rhox4* between 29.8Mb and 30.1Mb on the mouse X chromosome.

In order to validate the mouse genome sequence data, restriction analysis using the restriction endonucleases *AflIII, BstEII, EcoRV, HindIII, NdeI, ScaI, SspI* and *TthIII1,* which were predicted to reveal differences between duplicate copies, were performed (Table 4.1). Southern blotting, followed by analysis with a probe specific to *Rhox4* exon two, revealed the presence of all predicted bands, and band intensities apparently reflected the number of copies of each fragment expected for each digest pattern (Figure 4.1). These data therefore confirmed that there are seven copies of *Rhox4.* The sequences of all copies were extracted from RP23 BAC clones 43O20 (Accession number AL451076), 111C11 (AL589623), 167L6 (AL954686) and 457N24 (AL808133), which span the region, and were numbered *Rhox4.1* (closest to the centromere) to *Rhox4.7* (Figure 4.2). Analysis of the BAC sequences indicated that *Rhox4.1* to *4.5* are orientated 5' to 3' relative to the centromere, while duplicates *Rhox4.6* and *Rhox4.7* are in the opposite orientation.

To determine whether or not all *Rhox4* copies were likely to be expressed, and the extent of homology between the seven copies, the extracted genomic sequences were analysed in greater depth. The seven predicted *Rhox4* cDNA sequences all contain a start and stop codon and a polyadenylation signal, and are predicted to have the same

Table 4.1 Sizes of fragments predicted after digestion with restriction endonucleases for each *Rhox4* gene copy, when revealed with a probe to exon 2. No entry indicates that either a single or no restriction site is present from 4.2kb 5' and 2.5kb 3' of the copy.

	AflIII	BstEII	EcoRV	HindIII	NdeI	ScaI	SspI	TthIII1
Rhox4.1	3.1	10.0	1.4	5.2	9.6	9.1	4.1	
Rhox4.2	2.3		1.4	5.2	9.6	9.1	4.0	
Rhox4.3	4.0		1.4	25.7			5.9	
Rhox4.4	1.2	6.0	1.4		9.7		4.1	7.2
Rhox4.5	3.1	7.4	1.4	10.4	8.4		5.4	
Rhox4.6	3.1		1.4	13.6	8.3	9.1	4.1	
Rhox4.7	2.2		1.4	21.0	9.8	9.1	4.1	

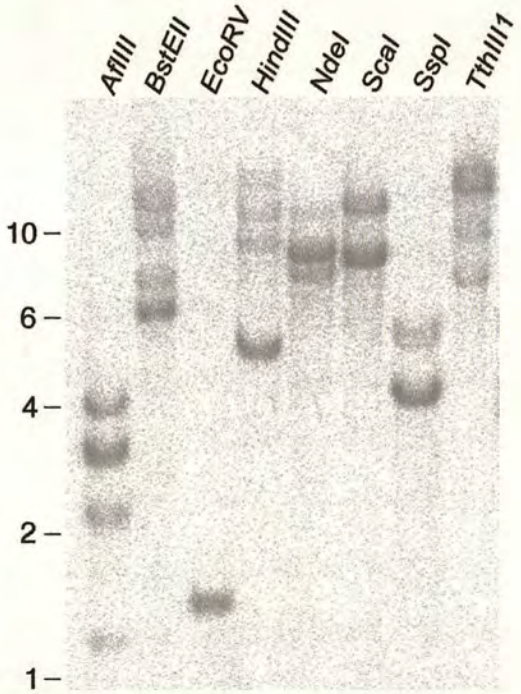


Figure 4.1 *Rhox4* exists as a seven-copy tandem array. C57BL/6 genomic DNA was digested with the restriction endonucleases shown and separated on a 0.8% agarose gel. After Southern blotting, a *Rhox4* exon 2 probe was used to reveal the restriction fragments. Multiple bands confirmed the presence of multiple gene copies, in agreement to predicted sizes. Position of size markers (kb) are indicated to the left of blot.

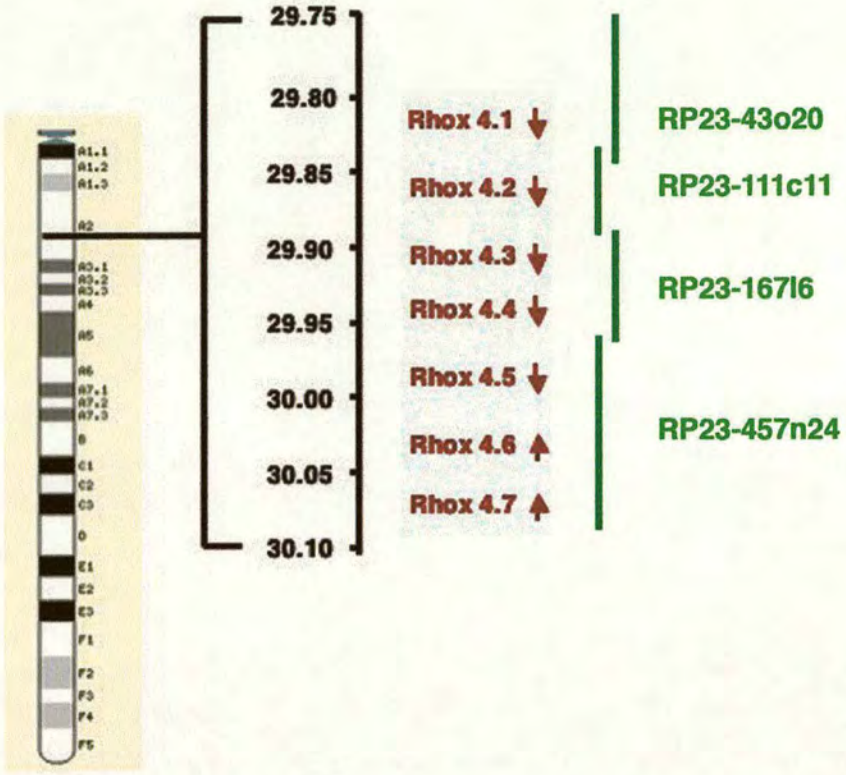


Figure 4.2 Position of the *Rhox4* seven-copy tandem array in the A2 region of mouse chromosome X. *Rhox4* gene copies were numbered *Rhox4.1* (closest to the centromere) to *Rhox4.7*. Arrows indicate gene orientation relative to the chromosome and the numbers are the position on the X chromosome relative to the centromere (Mb, NCBI build 33). 4 BAC clones span the region and are shown on the right of the diagram. X chromosome diagram taken from www.ensembl.org.

open reading frame. Multiple alignment of the cDNA sequences from the start codon to the polyadenylation codon showed that the sequences are highly conserved. The least degree of sequence identity was 98.2%, between copies *Rhox4.5* and *Rhox4.6* (Table 4.2a), and the highest was 99.7%, between *Rhox4.2* and *Rhox4.4;* mean identity was 99.0%. The region of the cDNA sequence encoding the homeodomain contained three polymorphic nucleotides and displayed comparative levels of sequence identity to the rest of the cDNA sequence (between 98.3% and 100% with a mean of 99.1%) (Table 4.2b). Inclusion of intronic sequences in the multiple alignment reduced the mean identity to 96.5%, with a range of 94.5% (between *Rhox4.3* and *Rhox4.7*) to 98.8% (*Rhox4.2* and *Rhox4.5*) (Table 4.2c), indicating that these regions are less highly conserved than the coding sequences. Multiple alignment of the 11.4kb region from 4.2kb 5' of the start codon to 2.5kb 3' of the polyadenylation signal resulted in a greater range of sequence identities (Table 4.2d). *Rhox4.3* and *Rhox4.7* displayed 88.6% identity whereas the same region of *Rhox4.1* and *Rhox4.4* was 98.5% identical; the mean identity was 93.8%.

The AVID global alignment program (Bray, Dubchak et al. 2003) was used to visualise the distribution of genomic sequence differences between the 11.4kb region of each *Rhox4* and the other six copies, representative plots with *Rhox4.1* and *Rhox4.6* as the base sequence are shown in Figure 4.3. This confirmed that, as described above, the coding regions were highly conserved, the greatest differences in the non-coding regions were located at the 3' end of the alignment, and that this 3' divergent region accounted for the lower degree of sequence identity between *Rhox4.3* and the other *Rhox4* copies (Table 4.2d). Interestingly, the other significant sequence differences were mainly concentrated in narrow regions 5' and 3' of the genes and within the introns (Figure 4.4). Some of these regions are due to a variable sequence present within the base sequence of *Rhox4.1* or *Rhox4.6*, identifiable by a variable region that occurs in one AVID global alignment but not the other. However, distinct regions that are polymorphic between all duplicates also exist. Further, examination of the sequence content of these regions (Figure 4.4a) reveals that they contain repeated sequences and small insertions and/or deletions indicating that they are regions of low sequence complexity.

Table 4.2 All seven *Rhox4* gene copies are highly conserved. Percentage identity and diversity of *Rhox4* copies within the a) cDNA (mean identity 99.0%); b) cDNA region encoding for the homeodomain (mean identity 99.1%); c) genomic sequence from the start codon to the polyadenylation signal (mean identity 96.5%); d) 11.4kb region from 4.2kb 5' of the ATG to 2.5kb 3' of the polyadenylation signal (mean identity 93.8%); e) peptide sequence (mean identity 97.3%); f) peptide sequence in homeodomain (mean identity 97.3%).

a)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		98.7	99.2	99.0	99.2	98.4	98.9	1	4-1 ORF
2	1.3		98.9	99.7	99.2	99.0	99.5	2	4-2 ORF
3	0.8	1.1		98.9	99.4	98.5	99.0	3	4-3 ORF
4	1.0	0.3	1.1		99.5	98.7	99.2	4	4-4 ORF
5	0.8	0.8	0.7	0.5		98.2	99.0	5	4-5 ORF
6	1.6	1.0	1.5	1.3	1.8		99.2	6	4-6 ORF
7	1.1	0.5	1.0	0.8	1.0	0.8		7	4-7 ORF
	1	2	3	4	5	6	7		

b)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		98.9	99.4	99.4	99.4	98.9	98.9	1	4-1 HB
2	1.1		98.3	99.4	98.3	100.0	100.0	2	4-2 HB
3	0.6	1.7		98.9	100.0	98.3	98.3	3	4-3 HB
4	0.6	0.6	1.1		98.9	99.4	99.4	4	4-4 HB
5	0.6	1.7	0.0	1.1		98.3	98.3	5	4-5 HB
6	1.1	0.0	1.7	0.6	1.7		100.0	6	4-6 HB
7	1.1	0.0	1.7	0.6	1.7	0.0		7	4-7 HB
	1	2	3	4	5	6	7		

c)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		98.1	96.4	98.7	97.8	97.5	95.4	1	4-1 + introns
2	1.6		95.5	97.6	98.8	96.9	94.6	2	4-2 + introns
3	3.6	3.8		96.0	95.8	96.0	94.5	3	4-3 + introns
4	1.2	1.6	3.6		97.7	97.3	95.1	4	4-4 + introns
5	2.0	1.1	3.6	1.7		96.7	94.6	5	4-5 + introns
6	2.1	2.2	3.6	2.2	2.3		95.6	6	4-6 + introns
7	3.7	4.0	4.5	3.8	3.9	3.7		7	4-7 + introns
	1	2	3	4	5	6	7		

d)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		97.3	90.7	98.5	95.2	97.0	93.8	1	4-1 12kb
2	1.8		90.3	97.2	95.4	96.4	93.1	2	4-2 12kb
3	9.0	9.0		90.7	90.0	90.1	88.6	3	4-3 12kb
4	1.1	1.9	8.8		95.0	96.8	93.9	4	4-4 12kb
5	4.0	3.5	9.2	4.2		94.2	92.3	5	4-5 12kb
6	2.6	2.6	9.7	2.7	5.0		93.9	6	4-6 12kb
7	5.0	5.1	10.4	5.0	6.6	5.1		7	4-7 12kb
	1	2	3	4	5	6	7		

e)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		96.6	98.5	97.6	98.1	95.6	97.1	1	Rhox4,1
2	3.5		97.1	99.0	97.6	97.1	98.5	2	Rhox4,2
3	1.5	3.0		97.1	98.5	96.1	97.6	3	Rhox4,3
4	2.5	1.0	3.0		98.5	96.1	97.6	4	Rhox4,4
5	2.0	2.5	1.5	1.5		94.7	97.1	5	Rhox4,5
6	4.5	3.0	4.0	4.0	5.5		97.6	6	Rhox4,6
7	3.0	1.5	2.5	2.5	3.0	2.5		7	Rhox4,7
	1	2	3	4	5	6	7		

f)

Percent Identity / Diversity

	1	2	3	4	5	6	7		
1		96.7	98.3	98.3	98.3	96.7	96.7	1	4-1 HD
2	3.4		95.0	98.3	95.0	100.0	100.0	2	4-2 HD
3	1.7	5.2		96.7	100.0	95.0	95.0	3	4-3 HD
4	1.7	1.7	3.4		96.7	98.3	98.3	4	4-4 HD
5	1.7	5.2	0.0	3.4		95.0	95.0	5	4-5 HD
6	3.4	0.0	5.2	1.7	5.2		100.0	6	4-6 HD
7	3.4	0.0	5.2	1.7	5.2	0.0		7	4-7 HD
	1	2	3	4	5	6	7		

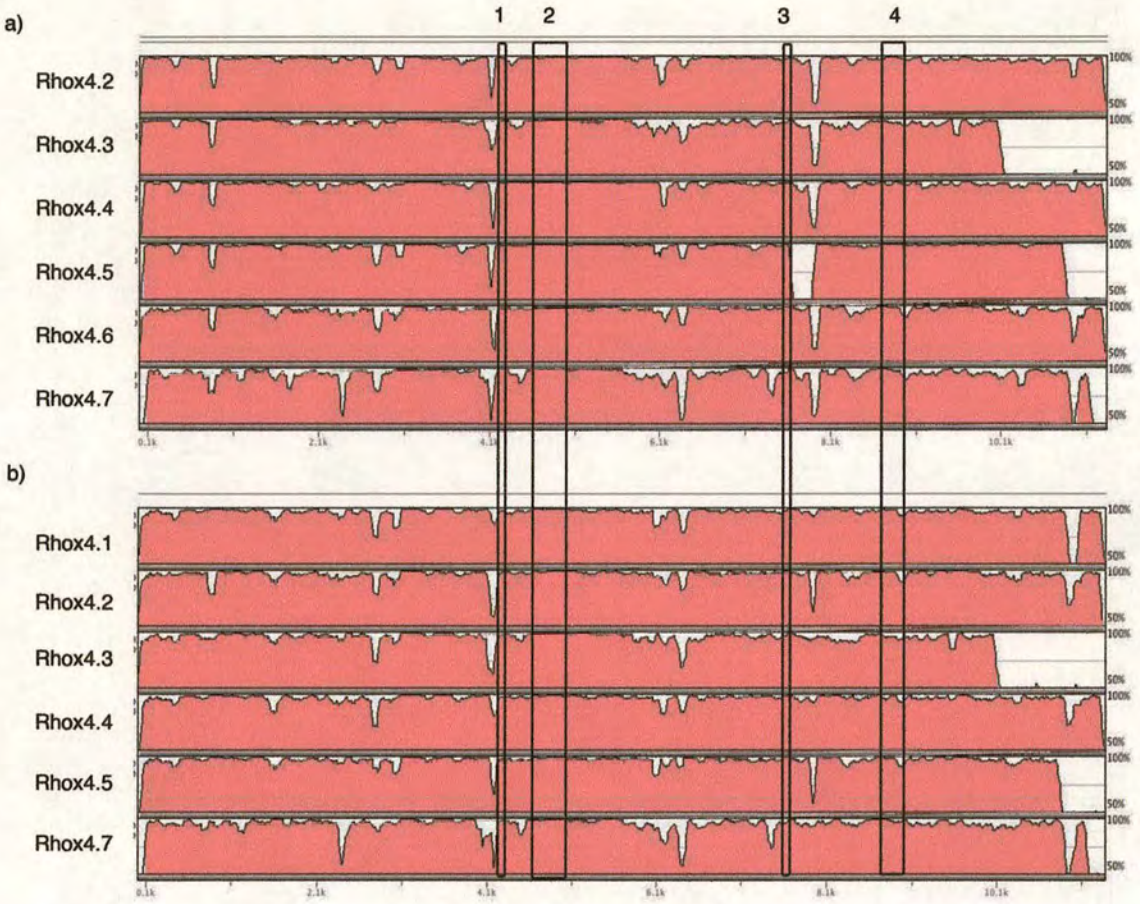
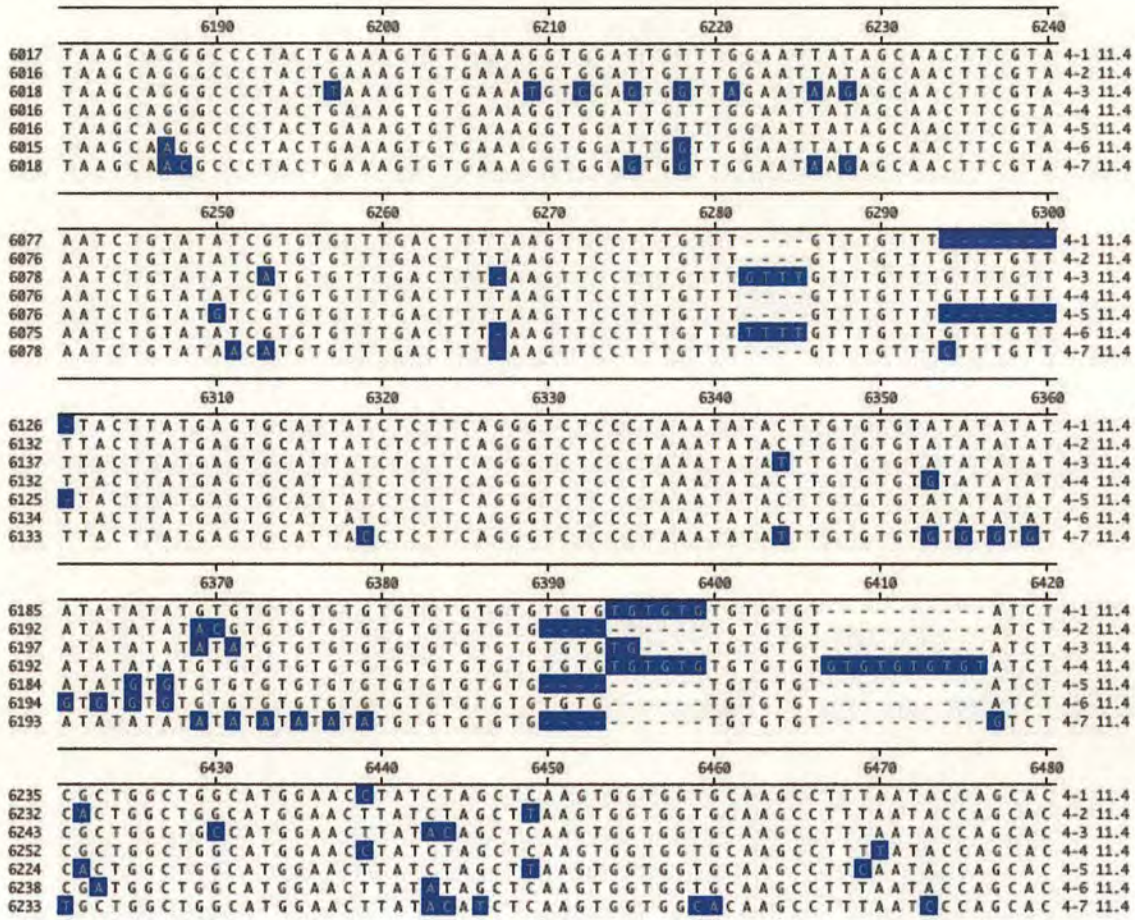


Figure 4.3 Genomic sequence differences between the seven *Rhox4* gene copies are found 3' of the gene, in distinct regions 5' of the gene and within the intronic sequences. AVID global alignment of an 11.4 kb region from 4.2kb 5' of the ATG to 2.5kb 3' of the polyadenylation signal of a) *Rhox4.1* aligned with *Rhox4.2* to *Rhox4.7* and b) *Rhox4.6* aligned with *Rhox4.1* to *Rhox4.5* and *Rhox4.7.* Percentage identity is shown on the vertical axis and position within the sequence on the horizontal axis. The position of the four exons are boxed and numbered.

a)



b)

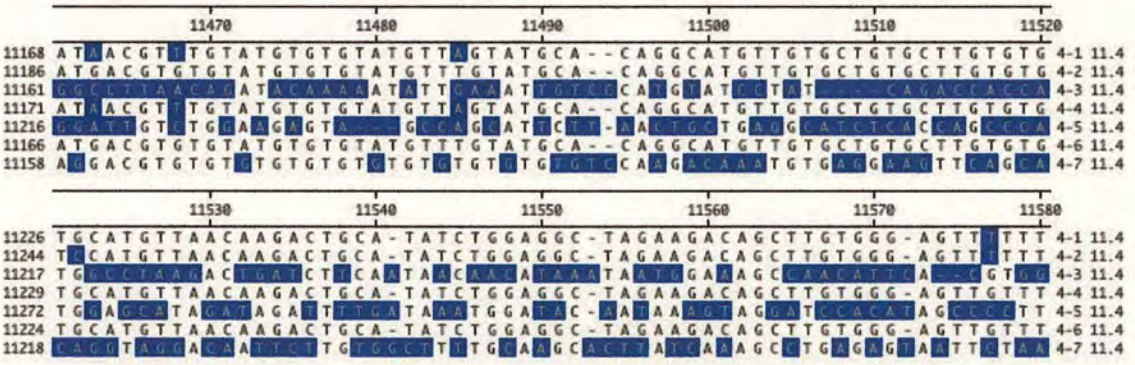


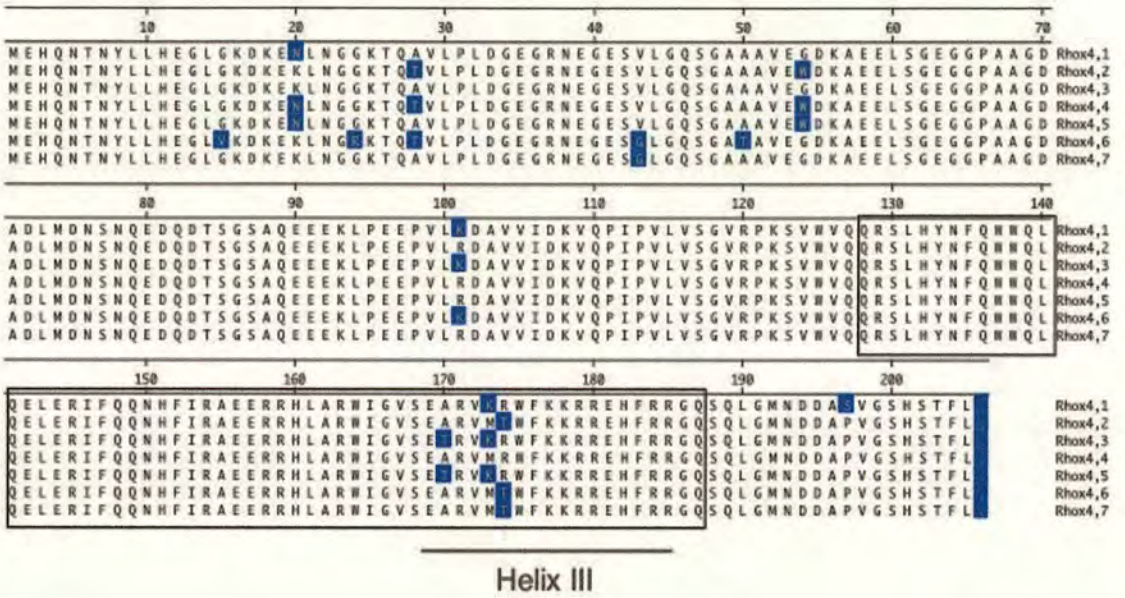
Figure 4.4 Examples of nucleic acid sequences containing significant sequence differences between *Rhox4* gene copies located in a) intron 2 and b) 2kb 3' of *Rhox4* exon 4. Sequence differences are highlighted. Numbers to left and above alignment indicate the position in an 11.4kb region of genomic DNA from 4.2kb 5' of the ATG to 4.5kb 3' of the polyadenylation signal.

Multiple alignment of the seven predicted *Rhox4* peptide sequences revealed a range of identities from 94.7% (between *Rhox4.5* and *Rhox4.6*) to 99.0% (between *Rhox4.2* and *Rhox4.4*) with a mean of 97.3% (Table 4.2e). Differences between the sequences were spread throughout the protein. The homeodomain also had a mean sequence identity of 97.3% (Table 4.2f), containing three polymorphic residues clustered in the third helix of the homeodomain (Figure 4.5). The arginine - threonine polymorphism that occurs in three *Rhox4* copies at position 47 of the homeodomain is of note, since this amino acid is involved in base-specific DNA contacts (Otting, Qian et al. 1990) and changes at this position may result in altered binding specificity.

The ratio of nonsynonymous substitutions per nonsynonymous site (*Ka*) to synonymous substitutions per synonymous site (*Ks*) is a measure of selective pressure. A ratio of one is indicative of neutral selection whilst a ratio greater than one suggests that positive evolutionary selection has been applied to the sequence. Inversely, a value of less than one is an indication of negative selection against changes in peptide sequence. The *Ka/Ks* ratio of the seven *Rhox4* copies is 2.69 when calculated using DnaSP (Rozas and Rozas 1999), suggesting that evolutionary mechanisms have acted on the locus in C57BL/6 inbred mice to enhance the number of nonsynonymous substitutions.

Pairwise alignment of the seven peptide sequences showed that eight of the twelve polymorphisms are present in multiple copies (Figure 4.5). Multiple rounds of mutation and duplication cannot easily explain the distribution of these polymorphisms, since the pattern of polymorphic residues between duplicates is complex. Therefore, the likely explanation for this observation is that the *Rhox4* locus has undergone multiple gene conversion events. Gene conversion is the non-reciprocal transfer of DNA sequence between homologous sequences (Orr-Weaver and Szostak 1985; Petes and Hill 1988), and is understood to be widespread between gene duplicates (Murti, Bumbulis et al. 1992; Hogstrand and Bohme 1997; Noonan, Grimwood et al. 2004). Examination of the intronic, upstream and downstream regions of the seven *Rhox4* copies showed that, due to the presence of particular

a)



b)

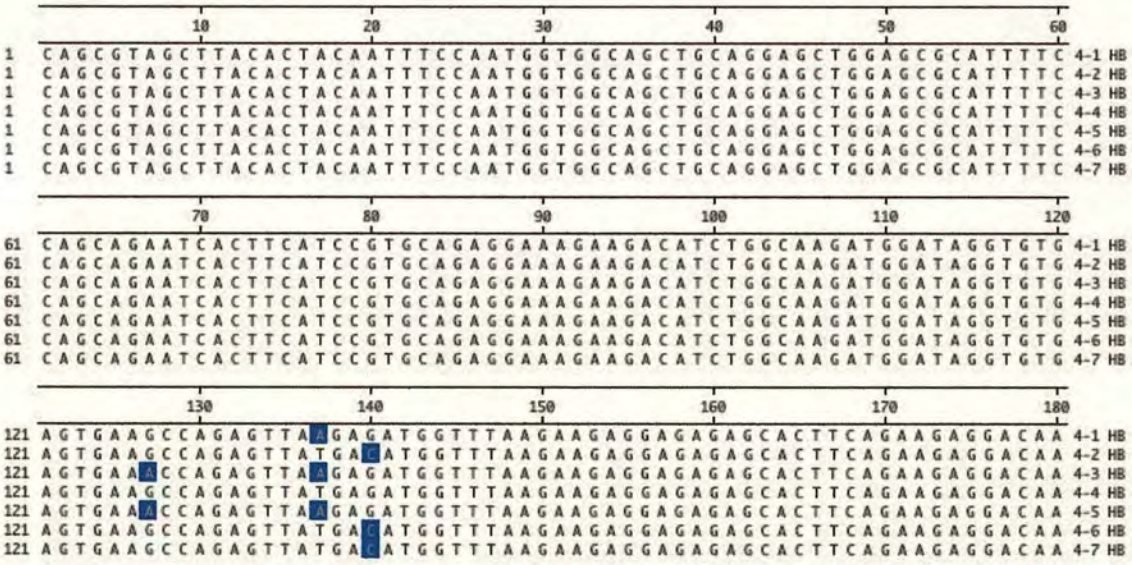


Figure 4.5 Eight out of twelve polymorphic residues are present in multiple Rhox4 gene copies. a) Pairwise alignment of the peptide sequences of all seven copies of Rhox4. Sequence differences are highlighted, including sequence corresponding to the 3rd helix of the homeobox domain. The boxed region contains the homeodomain. b) Pairwise alignment of the region of the cDNA sequence which encodes the homeodomain showing the three polymorphisms, which are present in multiple copies (highlighted).

sequence polymorphisms in multiple, non-adjacent copes, these regions are also likely to have undergone gene conversion (Figure 4.6). Whether the residues that have been converted, including in the homeodomain, result in altered protein function or expression pattern has yet to be determined.

The high level of sequence identity between the seven *Rhox4* copies observed in both the coding and non-coding sequences, suggest that the duplication events must be very recent. The analyses of the mouse locus described above were carried out on C57BL/6 genomic DNA. Therefore to determine whether the duplication of *Rhox4* was specific to C57BL/6 mice, locus organisation was investigated in the CBA and 129 inbred strains and in the wild sub-species *M. castaneous, M. spretus, M. caroli* and *M. pahari.* The phylogenetic relationship of the different sub-strains is shown in Figure 4.7a. Multiple bands were seen by Southern analysis in all sub-strains of mice investigated, including the non-musculus sub-strain, *Mus (Coelomys) pahari*, indicating the presence of more than one *Rhox4* copy (Figure 4.7b). Whether all sub-strains have seven copies as seen in C57BL/6 will require further investigation. However, this result indicates that the duplication events present within the *Rhox4* locus are likely to have taken place before the divergence of the above mouse sub-species, which may have occurred as recently as 5 million years ago (Bonhomme and Guenet 1996).

4.2.2 *Rhox2* and *3* are also present in the duplicated array

As for *Rhox4* (described above), a previous report described single copies of *Rhox1, 2* and *3* in the *Rhox* α cluster (Maclean, Chen et al. 2005). However, alignment of *Rhox2* and *Rhox3* but not *Rhox1* with the mouse genome sequence gave multiple hits, suggesting that *Rhox2* and *Rhox3* might also be present within the duplicated unit. Further examination of the alignments suggested eight copies of both *Rhox2* and *Rhox3*; the six most 5' copies of both genes being present 5' to 3' relative to the centromere, and the remaining two copies in the opposite orientation (Figure 4.8). In order to validate the mouse genome sequence data, restriction analysis were

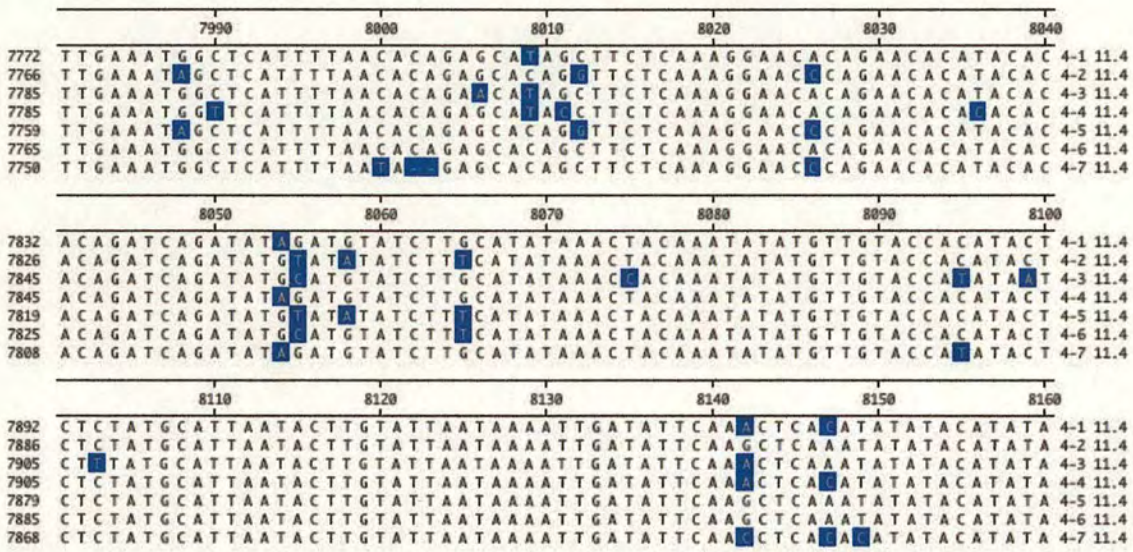
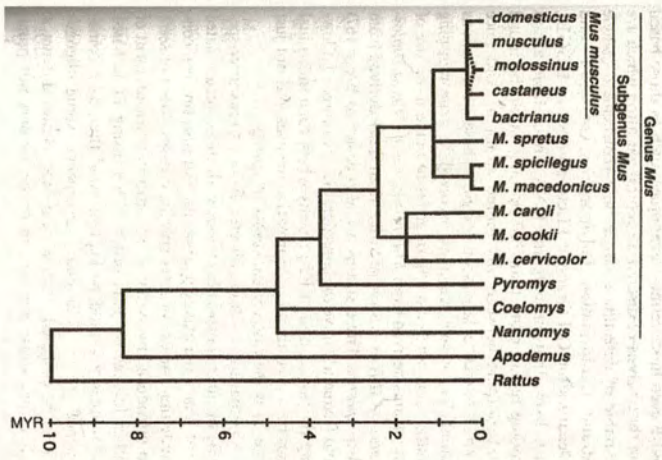


Figure 4.6 Gene conversion has occurred within the non-coding regions. Multiple alignment of genomic sequence from a region of intron 2 of the seven *Rhox4* gene copies, showing that polymorphic residues are found in multiple, non-adjacent copies. Sequence differences are highlighted. Numbers to left and above alignment indicate the position in an 11.4kb region of genomic DNA from 4.2kb 5' of the ATG to 4.5kb 3' of the polyadenylation signal.

a)



b)

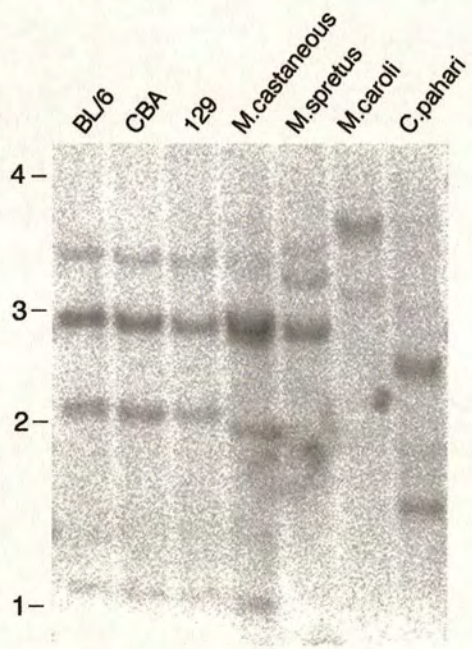


Figure 4.7 All sub-strains of mice investigated have multiple *Rhox4* gene copies. a) Phylogenetic tree of mouse sub-species taken from Bonhomme and Guenet (1996). b) Genomic DNA from three inbred strains of *M.(Mus) musculus domesticus*; C57BL/6, CBA and 129; and from *M.(Mus) musculus castaneous, M.(Mus) spretus, M.(Mus) caroli* and *M.(Coelomys) pahari* was digested with AflIII, separated on a 0.8% agarose gel, blotted and probed with a *Rhox4* exon 2 probe. Position of size markers (kb) are indicated to the left of blot.

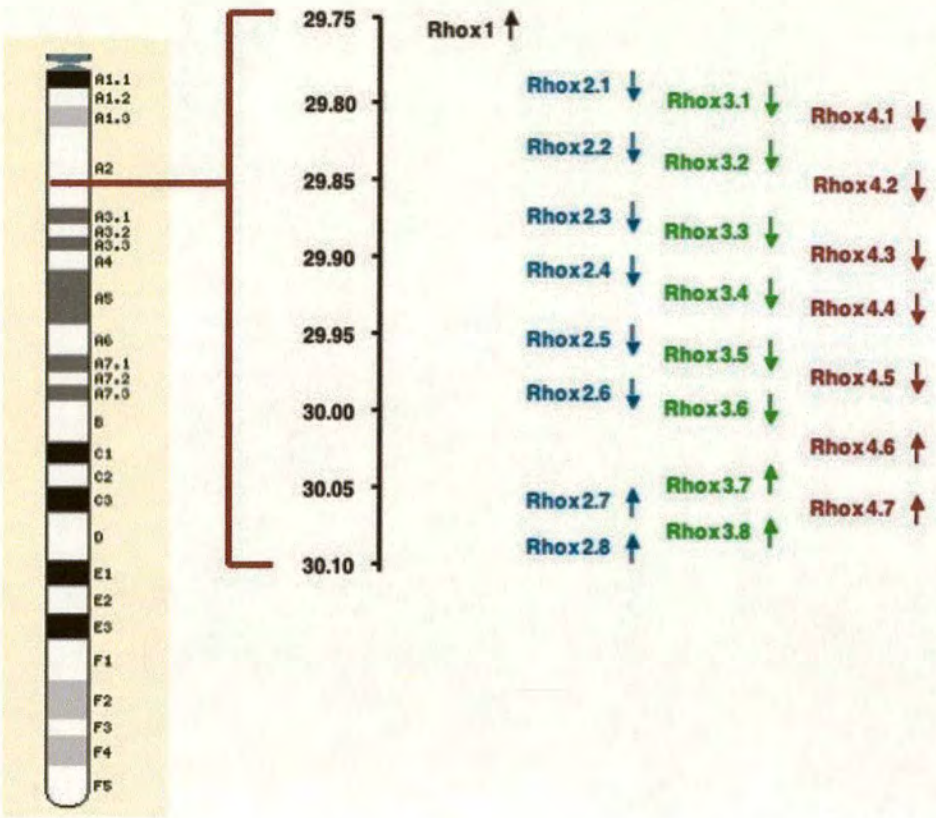


Figure 4.8 Position of the eight *Rhox2* and *3* gene copies and the seven *Rhox4* gene copies, which are present in a tandem array in the A2 region of mouse chromosome X. Copies were numbered *Rhox.1* (closest to the centromere) to *Rhox.7*. Arrows indicate gene orientation relative to the chromosome and the numbers are the position on the X chromosome relative to the centromere (Mb, NCBI build 33). X chromosome diagram taken from www.ensembl.org.

performed using restriction endonucleases, as described for *Rhox4.* Southern blotting followed by analysis with probes specific for either *Rhox2* or *Rhox3* confirmed the Band intensities apparently reflected the number of copies of each fragment expected for each digest pattern.

The predicted cDNA sequences of *Rhox2* and *Rhox3* were obtained from M.F. Wilkinson and J.A. Maclean (UTMD Anderson Cancer Centre, Houston, Texas), and the sequences of the eight copies were then extracted from the genomic sequence of four BACs that span the region (Figure 4.2). Comparison of these sequences showed that all copies of *Rhox2* and *3* were conserved. However, the degree of conservation of *Rhox2* and *Rhox3* copies was more variable than seen for *Rhox4* copies, as shown in Table 4.5 for *Rhox2* and *3* and Table 4.2 for *Rhox4. Rhox2* copy cDNA sequences were particularly variable, showing as low as 92.4% sequence identity (between copies *Rhox2.4* and *2.6*) (Table 4.5a). The minimum sequence identity for *Rhox3* copies was 97.0% (Table 4.5b) and 98.2% for *Rhox4* copies (Table 4.2a). Although the mean sequence identity between *Rhox2* copies was slightly lower than seen for *Rhox3* and *4* copies, the highest degree was 99.4%, between *Rhox2.2* and *2.5.* This range shows that *Rhox2* copies have undergone varying degrees of sequence divergence. Unlike *Rhox4*, variation in intron size was seen for both *Rhox2* and *Rhox3* copies. *Rhox2.5* intron 2 was 8.2kb, whereas all other copies had a 1.6kb intron 2, and intron 3 of copy *Rhox2.7* was only 1.8kb where all other copies had a 2.2kb intron. The only difference in intron size between *Rhox3* copies was found in copy *Rhox3.7,* where intron 2 was 2.7kb as opposed to 1.9kb. As expected, intronic, upstream and downstream regions of both *Rhox2* and *3* copies were less highly conserved than exon sequences (Table 4.5c,d). The mean identity of a region spanning 4.2kb 5' of the start codon to 2.5kb 3' of the polyadenylation signal was 80.5% for *Rhox2* copies and 89.3% for *Rhox3* copies, while the same region from *Rhox4* copies was 93.8% identical. Although the lower values obtained from *Rhox2* and *3* copies than for *Rhox4* copies were partially due to the differences in copy intron size, regions of significant sequence variation between copies were also present and these were visualised using the AVID global alignment program (Bray, Dubchak et al. 2003) (Figures 4.11 and 4.12). Figures contain representative

Table 4.3 Sizes of fragments (kb) predicted for each *Rhox2* gene copy after genomic DNA digested with restriction endonucleases shown is probed with an *Rhox2* exon 2 probe.

Rhox	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
AflIII	1.9	2.0	1.9	2.1	2.0	1.9	3.2	2.3
BstEII	>11	>11	>11	>11	9.4	>10	>14	>14
EcoRV	6.5	6.5	6.5	>8	5.4	6.5	>10	>10
HindIII	5.9	1.7	1.7	1.4	3.4	8.1	2.5	>8
NdeI	4.0	3.0	3.0	2.6	2.9	4.1	3.4	2.9
ScaI	5.5	5.6	5.6	6.1	3.5	5.6	6.1	6.0
SspI	1.6	1.6	1.6	1.6	5.5	1.6	1.9	1.9
TthIII1	>13	>13	>14	>14	>8	>14	>9	>14

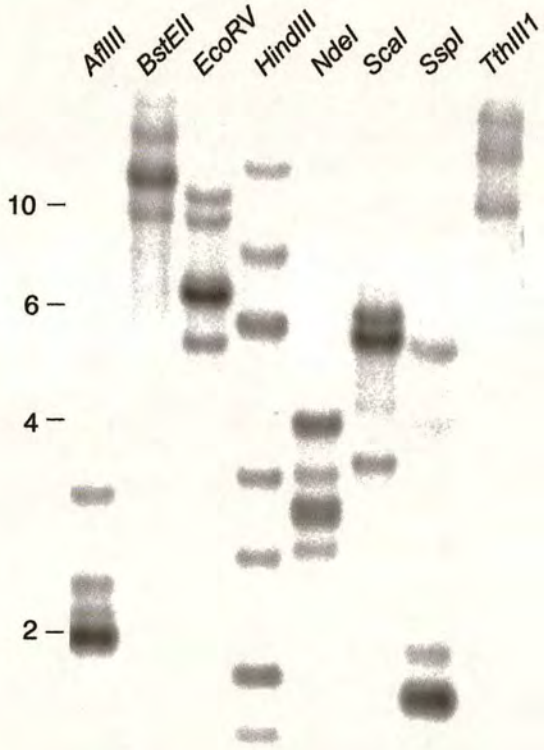


Figure 4.9 *Rhox2* exists as an eight-copy tandem array. C57BL/6 genomic DNA was digested with the restriction endonucleases shown and separated on a 0.8% agarose gel. After Southern blotting, a *Rhox2* probe was used to reveal the restriction fragments. Multiple bands confirmed the presence of multiple gene copies, in agreement with predicted sizes. Position of size markers (kb) are indicated to the left of blot.

Table 4.4 Sizes of fragments (kb) predicted for each *Rhox3* gene copy after Southern blotting of genomic DNA digested with restriction endonucleases shown, and probing with an *Rhox3* exon 2 probe.

Rhox	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8
AflIII	2.7	2.7	3.9	2.7	2.7	2.7	3.6	3.9
BstEII	0.8	0.8	>9	>9	0.8	0.8	0.8	>9
EcoRV	>12	11.8	>11	12.2	>12	>12	>13	>13
HindIII	12.3	11.0	>14	>11	11.0	>11	>11	2.9
NdeI	3.6	3.7	3.6	3.7	3.6	3.6	3.7	2.9
ScaI	1.7	1.7	2.4	1.7	1.7	1.7	1.7	0.7
SspI	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
TthIII1	3.6	3.6	3.6	>13	3.6	>14	>12	>11

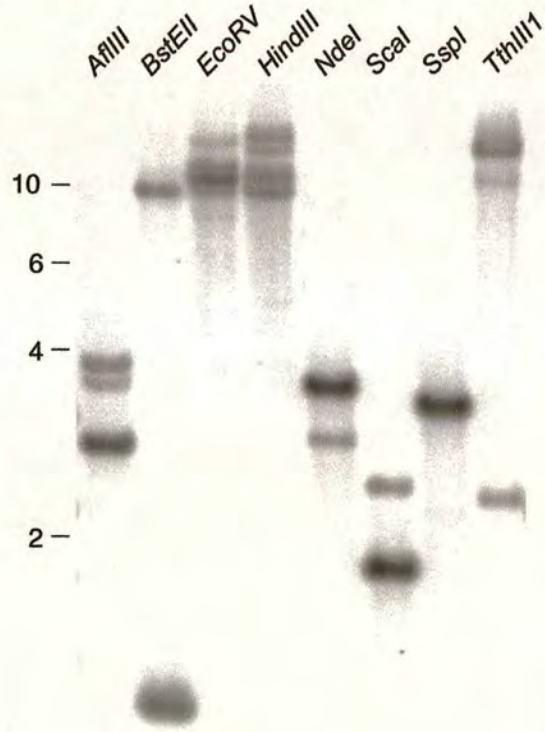


Figure 4.10 *Rhox3* exists as a eight-copy tandem array. C57BL/6 genomic DNA was digested with the restriction endonucleases shown and separated on a 0.8% agarose gel. After Southern blotting, a *Rhox3* probe was used to reveal the restriction fragments. Multiple bands confirmed the presence of multiple gene copies, in agreement with predicted sizes. Position of size markers (kb) are indicated to the left of blot.

Table 4.5 All eight *Rhox2* and *Rhox3* gene copies are conserved. Percentage identity and diversity of a) *Rhox2* copy cDNAs (mean identity 95.6%); b) *Rhox3* cDNAs (mean identity 98.3%); c) Region from 4.2kb 5' of *Rhox2* copy ATGs to 2.5kb 3' of the polyadenylation signal (mean identity 80.5%); d) Region from 4.2kb 5' of *Rhox3* copy ATGs to 2.5kb 3' of the polyadenylation signal (mean identity 89.3%); e) *Rhox2* peptide sequence (mean identity 93.8%); f) *Rhox3* peptide sequence (mean identity 92.9%); g) *Rhox2* homeodomain (mean identity 92.6%); h) *Rhox3* homeodomain (mean identity 99.2%).

a)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		96.8	96.8	96.1	97.2	95.1	96.5	95.5	1	2,1
2	3.3		99.0	93.6	99.4	97.6	94.0	92.9	2	2,2
3	3.3	1.0		94.0	99.1	97.3	94.6	93.3	3	2,3
4	4.0	6.7	6.3		93.8	92.4	97.7	95.6	4	2,4
5	2.9	0.6	0.9	6.4		97.9	94.2	93.2	5	2,5
6	5.1	2.5	2.7	8.0	2.1		92.8	93.7	6	2,6
7	3.6	6.3	5.6	2.4	6.0	7.6		95.5	7	2,7
8	4.6	7.4	7.0	4.5	7.1	6.6	4.6		8	2,8
	1	2	3	4	5	6	7	8		

b)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		99.3	98.9	99.6	99.9	97.7	97.7	98.3	1	3,1
2	0.7		98.4	99.4	99.3	97.2	97.5	98.3	2	3,2
3	1.1	1.6		98.8	98.8	97.3	97.1	98.2	3	3,3
4	0.4	0.6	1.2		99.6	97.6	97.6	98.4	4	3,4
5	0.1	0.7	1.2	0.4		97.7	97.7	98.3	5	3,5
6	2.3	2.8	2.7	2.5	2.3		98.3	97.2	6	3,6
7	2.3	2.6	3.0	2.5	2.3	1.7		97.0	7	3,7
8	1.7	1.7	1.8	1.6	1.7	2.8	3.1		8	3,8
	1	2	3	4	5	6	7	8		

c)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		98.4	98.0	80.2	98.4	96.6	69.7	80.1	1	Rhox2.1 11.4
2	1.2		98.9	79.0	98.5	97.3	68.4	78.8	2	Rhox2.2 11.4
3	1.5	0.9		78.6	98.3	97.1	68.2	78.4	3	Rhox2.3 11.4
4	16.9	17.2	17.4		78.9	78.4	69.8	84.0	4	Rhox2.4 11.4
5	1.3	1.4	1.4	17.4		61.4	43.4	49.8	5	Rhox2.5 18kb
6	2.8	2.3	2.5	17.7	2.6		68.2	78.4	6	Rhox2.6 11.4
7	31.4	32.0	31.9	31.7	31.9	32.5		78.7	7	Rhox2.7 11
8	17.2	17.5	17.7	16.2	17.7	17.9	22.9		8	Rhox2.8 11.4
	1	2	3	4	5	6	7	8		

d)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		98.0	84.3	90.9	99.3	94.3	90.2	89.7	1	3.1 10.8
2	1.0		84.2	91.3	98.0	94.4	90.8	90.4	2	3.2 10.8
3	15.9	16.2		81.2	84.3	81.4	81.0	82.9	3	3.3 10.8
4	3.9	4.3	15.5		90.9	90.4	91.4	86.6	4	3.4 10.8
5	0.5	1.0	15.9	4.0		94.4	90.3	89.9	5	3.5 10.8
6	4.6	4.6	19.5	4.9	4.6		92.6	87.9	6	3.6 10.8
7	8.2	7.7	19.2	3.2	8.3	5.9		80.3	7	3.7 11.6
8	9.6	9.2	17.6	9.2	9.5	11.8	12.4		8	3.8 10.8
	1	2	3	4	5	6	7	8		

e)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		97.9	97.9	90.6	100.0	94.8	92.7	93.8	1	2,1I
2	2.1		97.9	90.1	97.9	93.8	92.2	93.2	2	2,2I
3	2.1	2.1		90.6	97.9	93.8	93.8	94.8	3	2,3I
4	10.0	10.6	10.0		90.6	87.5	92.7	93.8	4	2,4I
5	0.0	2.1	2.1	10.0		94.8	92.7	93.8	5	2,5I
6	5.4	6.5	6.5	13.7	5.4		88.5	89.6	6	2,6I
7	7.7	8.3	6.5	7.7	7.7	12.5		99.0	7	2,7
8	6.5	7.1	5.4	6.5	6.5	11.2	1.0		8	2,7I
	1	2	3	4	5	6	7	8		

f)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		74.3	98.6	99.5	100.0	96.7	74.3	98.6	1	3,1P
2	1.3		97.5	98.1	98.8	95.7	97.5	97.5	2	3,2P
3	1.4	2.5		98.1	98.6	97.2	74.3	98.6	3	3,3P
4	0.5	1.9	1.9		99.5	96.3	73.8	98.1	4	3,4P
5	0.0	1.3	1.4	0.5		96.7	74.3	98.6	5	3,5P
6	3.3	4.5	2.9	3.8	3.3		72.9	97.7	6	3,6P
7	1.3	2.5	1.3	1.9	1.3	3.2		98.8	7	3,7P
8	1.4	2.5	1.4	1.9	1.4	2.4	1.3		8	3,8P
	1	2	3	4	5	6	7	8		

g)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		100.0	100.0	90.0	100.0	95.0	91.7	90.0	1	2.1 HD
2	0.0		100.0	90.0	100.0	95.0	91.7	90.0	2	2.2 HD
3	0.0	0.0		90.0	100.0	95.0	91.7	90.0	3	2.3 HD
4	10.8	10.8	10.8		90.0	86.7	88.3	85.0	4	2.4 HD
5	0.0	0.0	0.0	10.8		95.0	91.7	90.0	5	2.5 HD
6	5.2	5.2	5.2	14.7	5.2		86.7	95.0	6	2.6 HD
7	8.9	8.9	8.9	12.7	8.9	14.7		85.0	7	2.7 HD
8	10.8	10.8	10.8	16.8	10.8	5.2	16.8		8	2.8 HD
	1	2	3	4	5	6	7	8		

h)

Percent Identity / Diversity

	1	2	3	4	5	6	7	8		
1		100.0	100.0	98.3	100.0	98.3	100.0	100.0	1	3.1 HD
2	0.0		100.0	98.3	100.0	98.3	100.0	100.0	2	3.2 HD
3	0.0	0.0		98.3	100.0	98.3	100.0	100.0	3	3.3 HD
4	1.7	1.7	1.7		98.3	96.7	98.3	98.3	4	3.4 HD
5	0.0	0.0	0.0	1.7		98.3	100.0	100.0	5	3.5 HD
6	1.7	1.7	1.7	3.4	1.7		98.3	98.3	6	3.6 HD
7	0.0	0.0	0.0	1.7	0.0	1.7		100.0	7	3.7 HD
8	0.0	0.0	0.0	1.7	0.0	1.7	0.0		8	3.8 HD
	1	2	3	4	5	6	7	8		

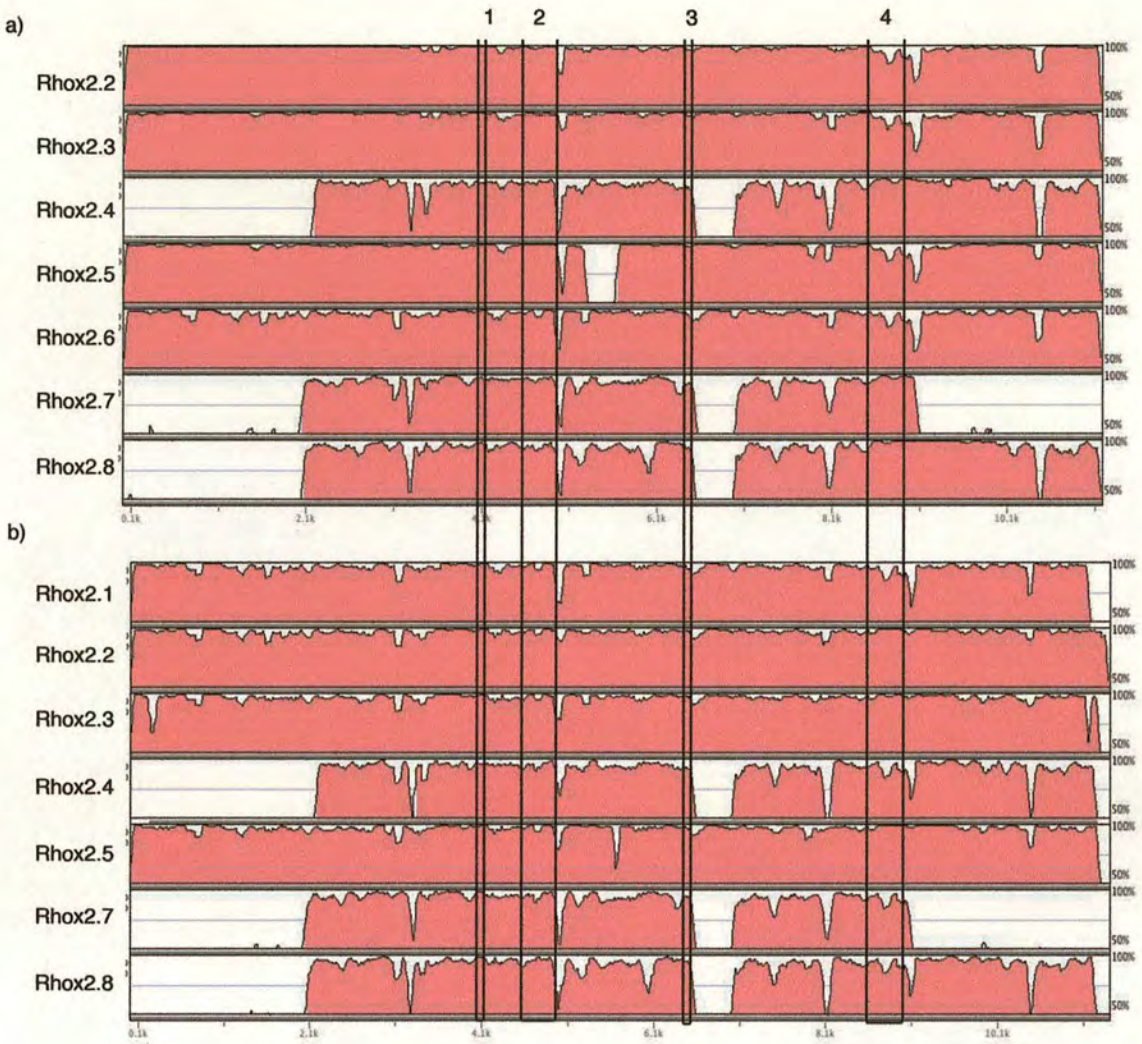


Figure 4.11 Genomic sequence differences between the eight *Rhox2* gene copies are found in distinct regions 5' and 3' of the gene and within the intronic sequences. AVID global alignment of a region spanning 4.2kb 5' of the ATG to 2.5kb 3' of the polyadenylation signal of a) *Rhox2.1* aligned with *Rhox2.2* to *Rhox2.7* and b) *Rhox2.6* aligned with *Rhox2.1* to *Rhox2.5* and *Rhox2.7*. Percentage identity is shown on the vertical axis and position within the sequence on the horizontal axis. The position of the four exons are boxed and numbered.

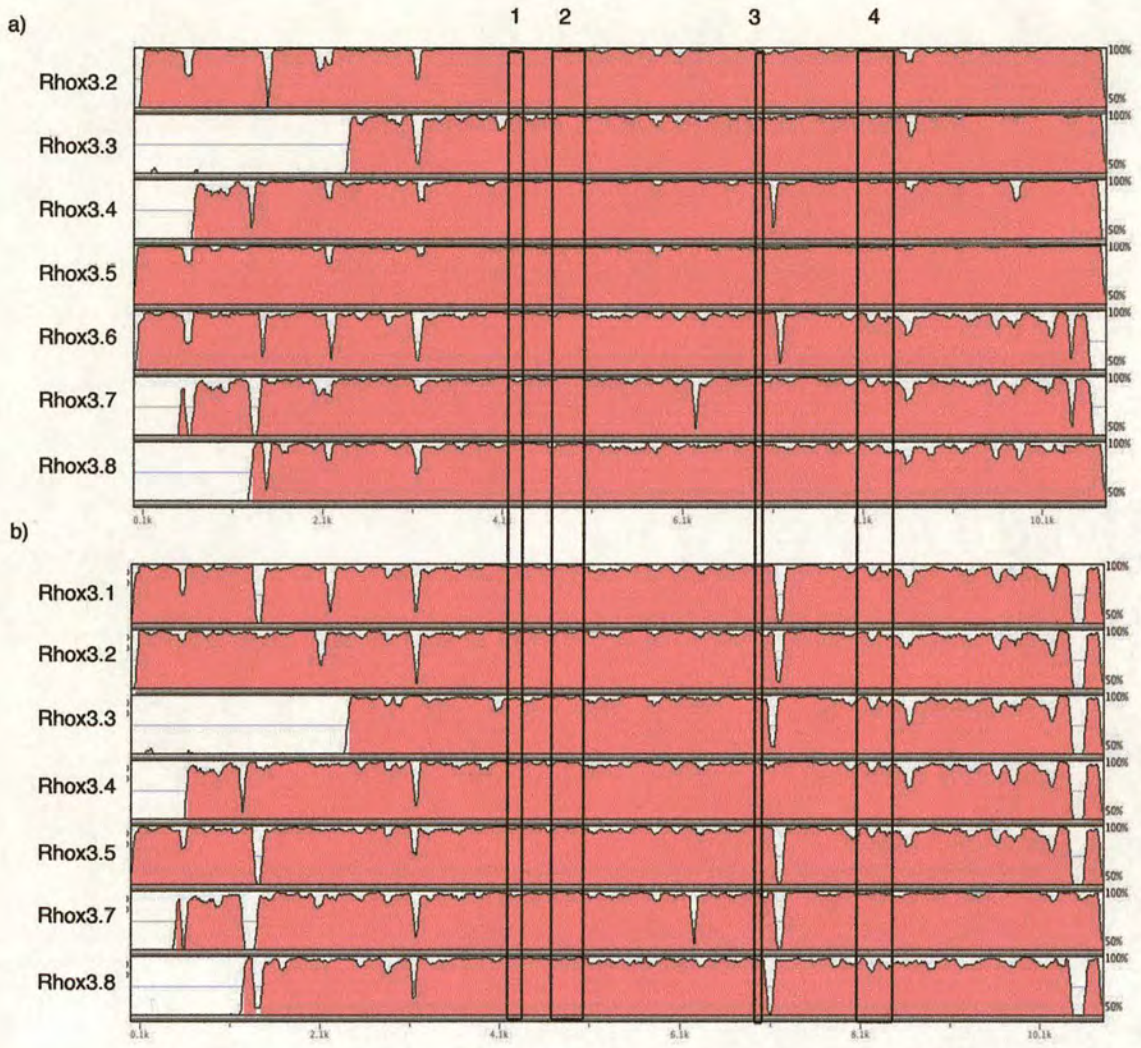


Figure 4.12 Genomic sequence differences between the eight *Rhox3* gene copies are found in distinct regions 5' and 3' of the gene and within the intronic sequences. AVID global alignment of a region spanning 4.2kb 5' of the ATG to 2.5kb 3' of the polyadenylation signal of a) *Rhox3.1* aligned with *Rhox3.2* to *Rhox3.7* and b) *Rhox3.6* aligned with *Rhox3.1* to *Rhox3.5* and *Rhox3.7*. Percentage identity is shown on the vertical axis and position within the sequence on the horizontal axis. The position of the four exons are boxed and numbered.

alignments, each showing plots of the first and sixth copies aligned with seven other copies for both *Rhox2* and *Rhox3*. Two alignments are shown to distinguish between sequences which are divergent only in the base sequence and sequences which are divergent between multiple copies. Differences in *Rhox2* copies were found in large regions both 5' and 3' of the gene and within intron three and, as seen for *Rhox4* copies, there were also small, distinct regions which were divergent between copies (Figure 4.11). Differences between *Rhox3* copies were found primarily 5' of the gene, although again short regions of sequence difference were found in introns and 5' and 3' regions (Figure 4.12).

Multiple alignment of the eight *Rhox2* copies predicted protein sequences showed that there are thirty three polymorphic residues, and the identity of copies varied from only 87.5% (between *Rhox2.4* and *2.6*) to 100% (between *Rhox2.1* and *2.5*) (Table 4.5e). Eleven polymorphic residues were present within the homeodomain (Figure 4.13) and were concentrated at the N-terminus and within the second helix. Unlike *Rhox4,* no polymorphisms were present in residues with established roles in helix packing or DNA specificity. However the glutamine – lysine polymorphism found in *Rhox2.7* at position 42 of the homeodomain is in a residue that is conserved between all *Rhox* family members except *Rhox5* (Maclean, Chen et al. 2005), therefore suggesting that the function of this copy may be affected by the presence of this polymorphism.

Translation of *Rhox3* gene copy cDNA sequences predicted copies *3.2* and *3.7* to have a peptide sequence that contained only 51 residues (Figure 4.14), due to the presence of a stop codon 30 nucleotides after the first ATG. Since a second, down-stream start codon is present in the same reading frame, all copies could theoretically generate a truncated protein missing the N-terminus. Either start codon could be utilised by *Rhox3* copies lacking the stop codon. However, examination of the Kozak sequences surrounding the two ATGs showed that the second but not the first ATG contained a G at position +1, which is thought to be important for efficient translation (Kozak 1986) and is conserved in all *Rhox* proteins except *Rhox11*. The

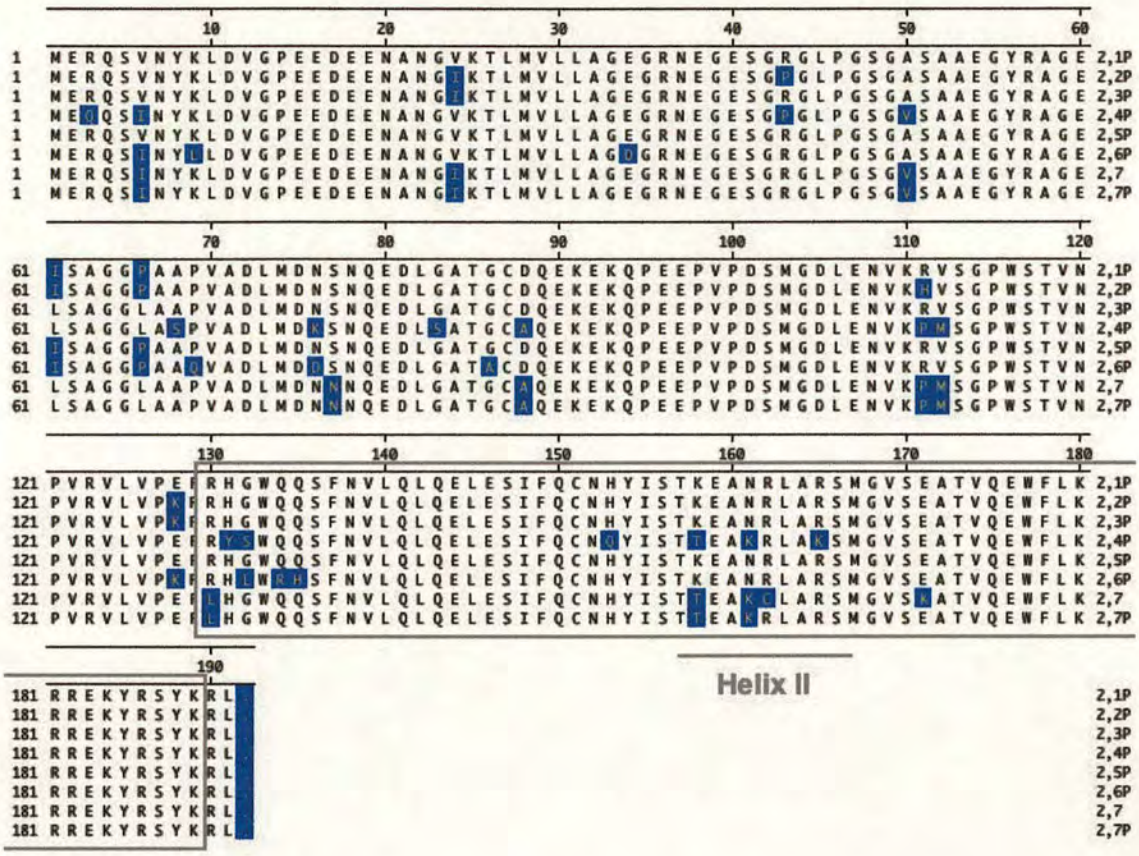


Figure 4.13 There are thirty polymorphic residues between the eight *Rhox2* gene copies. Multiple alignment of the peptide sequences of all eight copies of *Rhox2.* Sequence differences are highlighted, including sequence corresponding to the 2rd helix of the homeobox domain, the boxed region contains the homeodomain.

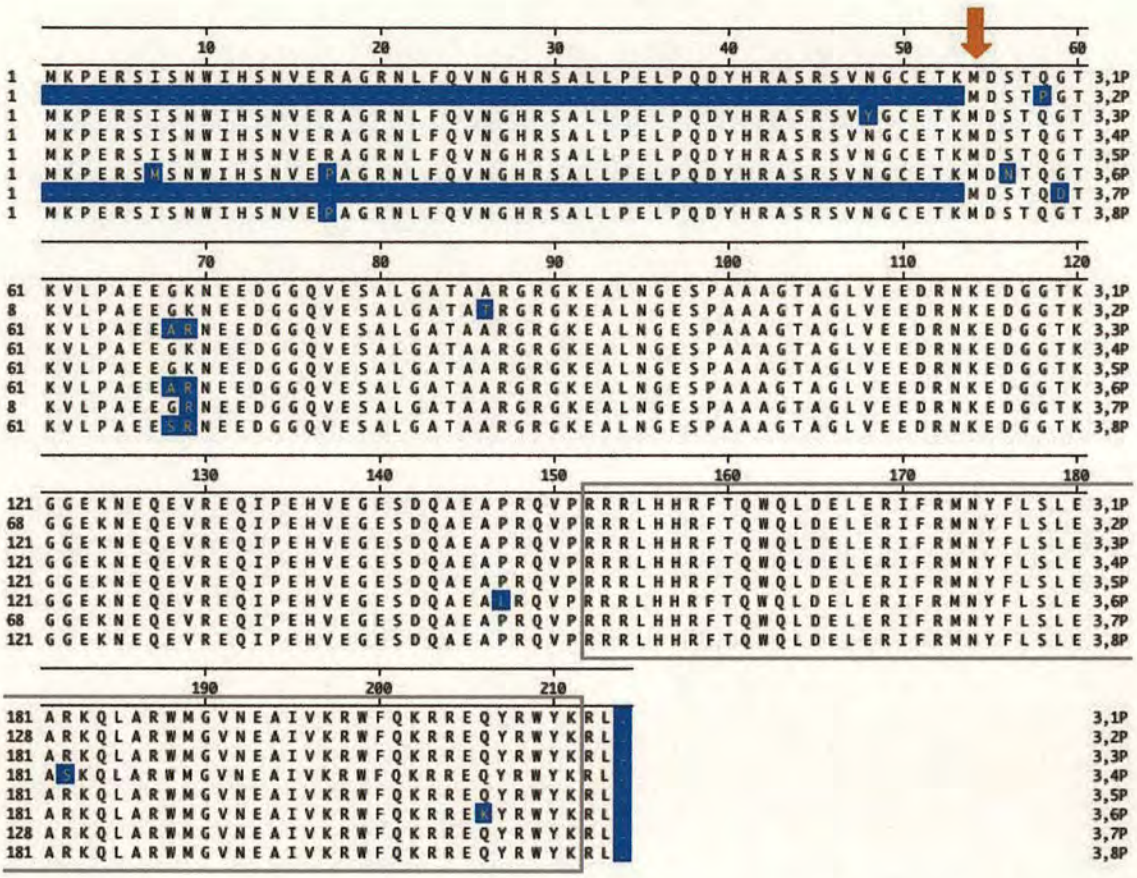


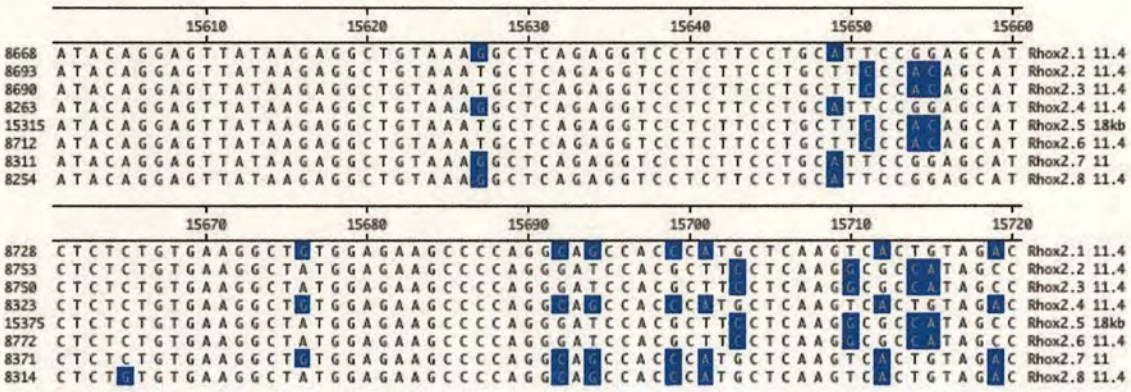
Figure 4.14 There are two possible start codons within six of the eight *Rhox3* gene copies. Multiple alignment of the predicted peptide sequences of all eight copies of *Rhox3.* Copies *Rhox3.1, 3.3-3.6* and *3.8* may transcribe a product that is 53 residues longer than *Rhox3.2* and *3.7.* The arrow indicates the position of the downstream methionine. Sequence differences are highlighted and the homeodomain is contained within the box.

presence of an upstream open reading frame within *Rhox3.2* and *3.7* is likely to reduce translation efficiency from these copies. Utilisation of the second start codon would result in a product of only 160 amino acids in length, which would make *Rhox3* the shortest *Rhox* family member. *Rhox12* is currently the shortest at 187 residues and the average is 222 residues. In peptide sequence identity calculations for *Rhox3,* it was assumed that the first start codon is utilised by all copies except *Rhox3.2* and *3.7* and that two different sized products are present (Table 4.5f).

With the exception of the difference in product size, very few sequence polymorphisms are present between *Rhox3* copies (Figure 4.14). Two polymorphisms are found in the homeodomain, again neither are present in residues with roles in helix packing or DNA binding specificity. However the polymorphism in *Rhox3.4* at position 31 is in a residue conserved in all *Rhox* family members except *Rhox2* and *Rhox7* (Maclean, Chen et al. 2005).

To determine whether selective pressure appeared to have operated in the *Rhox2* or *Rhox3* copies, the *Ka/Ks* ratios between copies at each locus were examined. Although less than 1, the *Ka/Ks* ratios for *Rhox2* and *Rhox3* copies were also very high for gene duplicates, being 0.67 and 0.53 respectively, again suggesting that positive selection or gene conversion events may have occurred. These values are much lower than the *Ka/Ks* ratio for *Rhox4* (2.69). The presence of the same sequence polymorphisms in multiple, non-adjacent copies of both the protein sequence of *Rhox4* and intron, upstream and downstream sequence suggested that multiple gene conversion events had taken place. Examination of peptide and genomic sequences of *Rhox2* and *Rhox3* copies showed that gene conversion events are also likely to have taken place at these loci (see Figure 4.13, 4.14 and 4.15). The difference in *Ka/Ks* ratio between *Rhox4* copies and *Rhox2* and *3* copies is due to the very high proportion (67%) of *Rhox4* nonsynonymous polymorphisms that are present in more than one copy. Only 42% of *Rhox2* and 23% of *Rhox3* peptide sequence polymorphisms are found in more than one copy, suggesting that either increased numbers of gene conversion events may have occurred between *Rhox4*

a)



b)

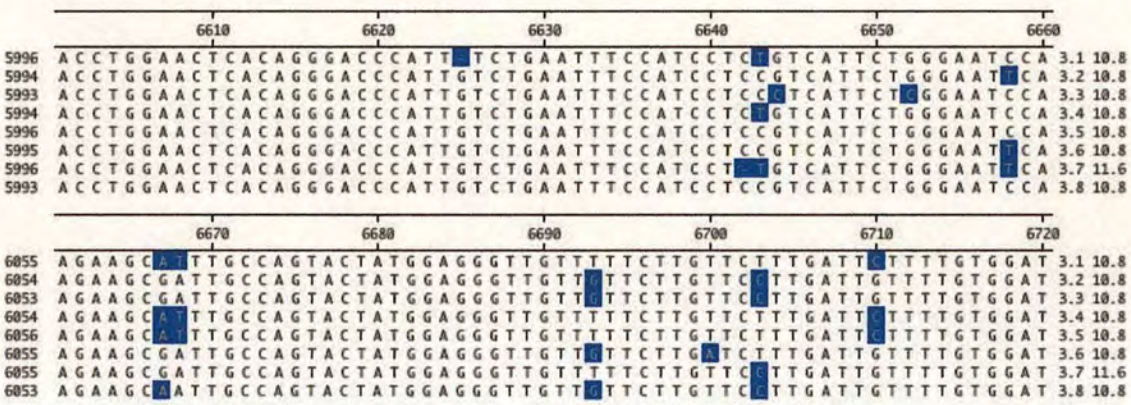


Figure 4.15 Gene conversion events have occurred within the non-coding regions of *Rhox2* and *Rhox3* gene copies. Multiple alignment of the genomic sequence from a) a region of intron 3 of the eight *Rhox2* copies and b) a region of intron 2 of the eight *Rhox3* copies, showing that polymorphic residues are found in multiple, non-adjacent copies. Sequence differences are highlighted. Numbers to the left of alignment indicate the position of each sequence in the alignment which contains from 4.2kb 5' of the ATG to 2.5kb 3' of the polyadenylation signal. Numbers above plot indicate the position within the whole alignment.

copies compared to the *Rhox2* or *3* copies, or that increased selective pressure for advantageous mutations has been applied at the *Rhox4* copies.

4.2.3 Identification of Rat Orthologues of *Rhox3* and *Rhox4*

The presence of multiple copies of *Rhox4* in all sub-species of mouse investigated suggests that the duplication events present in the *Rhox* α cluster took place before their divergence. I was therefore interested in identifying orthologues of the *Rhox* α cluster genes to determine when the duplication events occurred. The syntenic region of the rat genome sequence (Xq11), although currently incomplete, contains two predicted genes that display a high degree of similarity to genes present in the mouse *Rhox* α cluster, including the characteristic arrangement of the homeobox-containing region between three exons. *Rattus norvegicus similar to homeobox protein EHOX* (XM_233314) displays the highest level of identity to *Rhox4* (63.1% amino acid sequence, 77.8% nucleotide across the open reading frame; Table 4.6a,b and Figure 4.16) and *Rattus norvegicus similar to RIKEN cDNA 4930539I12* (XM_343759) displays the highest level of identity to *Rhox3* (64.0% amino acid sequence, 83.2% nucleotide, Table 4.6a,b and Figure 4.16), suggesting that these are rat orthologues of mouse *Rhox4* and *Rhox3*. The level of peptide sequence identity of both sets of predicted orthologues was significantly greater within the homeodomain than over the whole peptide sequence; 83.3% sequence identity for *Rhox3* orthologues within the homeodomain and 76.7% for *Rhox4* orthologues (Table 4.6c and Figure 4.16).

The level of identity between mouse and rat for both genes is very low, since on average rat and mouse genes are 94% identical at the amino acid level (Wolfe and Sharp 1993). However, mouse *Rhox3* and *Rhox4* are more identical in both cDNA and peptide sequence to their predicted rat orthologues than they are to all other mouse *Rhox* family members (Table 4.6e,f), suggesting that these orthologue predictions are correct. Furthermore, some sequence identity was also present between the genomic sequences of mouse and rat orthologues of *Rhox3* and *4*. *Rhox3*

Table 4.6 Identities of predicted mouse and rat *Rhox3* and *Rhox4* orthologues. Percentage identity and divergence of a) cDNA sequences; b) protein sequences; c) homeodomain sequences d) genomic sequence from 4.2kb 5' of start codon to 2.5kb 3' of polyadenylation signal e) mouse *Rhox* family member's cDNA sequences and f) mouse *Rhox* family member's protein sequences.

a)

Percent Identity

Diversity	1	2	3	4		
1		40.1	50.4	83.2	1	Mouse Rhox3.1
2	64.9		77.8	55.3	2	Mouse Rhox4.1
3	62.6	26.0		50.8	3	Rat Rhox4 predicted
4	17.3	60.0	59.8		4	Rat Rhox3 predicted
	1	2	3	4		

b)

Percent Identity

Diversity	1	2	3	4		
1		29.0	30.4	64.0	1	Mouse Rhox3.1
2	134.4		63.1	32.5	2	Mouse Rhox4.1
3	126.6	39.7		36.3	3	Rat Rhox4 predicted
4	47.1	120.0	106.9		4	Rat Rhox3 predicted
	1	2	3	4		

c)

Percent Identity

Diversity	1	2	3	4		
1		83.3	58.3	78.3	1	3.1 HD
2	18.9		56.7	76.7	2	Rat Rhox3 HD
3	60.0	63.7		76.7	3	4-1 HD
4	25.6	28.0	28.0		4	Rat Rhox4 HD
	1	2	3	4		

d)

Percent Identity

Diversity	1	2	3	4		
1		67.1	39.2	36.9	1	Mouse Rhox3.1
2	38.0		38.7	36.4	2	Rat Rhox3 predicted
3	118.4	121.5		52.8	3	Mouse Rhox4.1
4	133.1	135.8	70.0		4	Rat Rhox4 predicted
	1	2	3	4		

e)

Percent Identity

Diversity	1	2	3	4	5	6	7	8	9	10	11	12		
1		50.0	44.5	57.5	40.1	39.9	42.2	42.9	40.4	24.1	32.5	38.0	1	Rhox 1 cDNA
2	47.1		54.4	64.9	42.9	45.4	40.7	45.6	43.4	30.6	37.5	37.7	2	2,1
3	80.7	64.6		50.8	42.7	44.4	37.5	45.2	43.8	28.4	36.8	36.3	3	3,1
4	53.7	45.5	79.6		39.8	41.8	40.9	43.3	40.3	27.4	33.8	36.9	4	Rhox4-1
5	98.0	92.7	100.1	101.3		47.7	31.5	40.3	47.4	25.2	32.7	31.4	5	Rhox 5 cDNA
6	105.1	87.5	98.5	102.1	83.4		33.6	44.5	88.3	29.1	36.9	36.8	6	Rhox 6 cDNA
7	108.9	88.1	100.6	89.9	133.5	124.1		39.1	31.3	21.3	27.1	29.2	7	Rhox 7 cDNA
8	106.3	85.1	89.5	92.0	107.5	96.0	120.0		33.2	19.8	25.9	28.0	8	Rhox 8 cDNA
9	118.2	88.0	94.2	103.6	80.4	9.2	125.3	95.5		26.4	34.4	35.3	9	Rhox 9 cDNA
10	122.2	118.6	121.3	111.1	133.8	107.3	118.3	117.4	110.8		54.7	41.6	10	Rhox 10 cDNA
11	124.7	121.7	125.6	126.5	141.0	116.7	125.0	121.8	117.5	71.0		37.2	11	Rhox 11 cDNA
12	126.9	125.5	136.2	127.7	163.5	131.4	154.6	124.9	135.6	113.1	128.5		12	Rhox 12 cDNA
	1	2	3	4	5	6	7	8	9	10	11	12		

f)

Percent Identity

Diversity	1	2	3	4	5	6	7	8	9	10	11	12		
1		26.2	21.3	23.6	16.0	18.7	21.8	19.6	19.1	14.7	14.2	17.3	1	Rhox 1
2	135.0		30.2	51.6	18.2	22.4	38.0	26.6	23.4	16.7	14.1	16.7	2	2,1P
3	209.0	145.8		29.4	17.8	21.5	22.9	25.2	22.4	12.6	14.0	18.2	3	3,1P
4	153.8	73.0	130.8		19.4	22.8	40.8	26.2	25.2	16.0	13.6	17.5	4	Rhox4,1
5	233.0	243.0	229.0	231.0		20.9	18.5	24.2	21.3	12.8	13.3	13.7	5	Rhox 5
6	221.0	215.0	209.0	210.0	210.0		21.1	24.6	82.0	14.5	14.5	15.8	6	Rhox 6
7	202.0	116.8	210.0	107.1	249.0	229.0		16.3	17.4	13.8	9.6	13.8	7	Rhox 7
8	249.0	182.4	194.1	168.4	166.7	184.8	276.0		19.5	13.0	15.4	13.3	8	Rhox 8
9	216.0	205.0	200.0	191.6	205.0	20.6	224.0	181.0		14.0	14.0	17.1	9	Rhox 9
10	221.0	214.0	271.0	231.0	315.0	244.0	213.0	190.7	252.0		33.5	21.8	10	Rhox 10
11	230.0	261.0	249.0	277.0	311.0	249.0	315.0	159.9	257.0	138.3		20.3	11	Rhox 11
12	209.0	243.0	213.0	236.0	298.0	248.0	234.0	213.0	228.0	186.2	195.6		12	Rhox 12
	1	2	3	4	5	6	7	8	9	10	11	12		

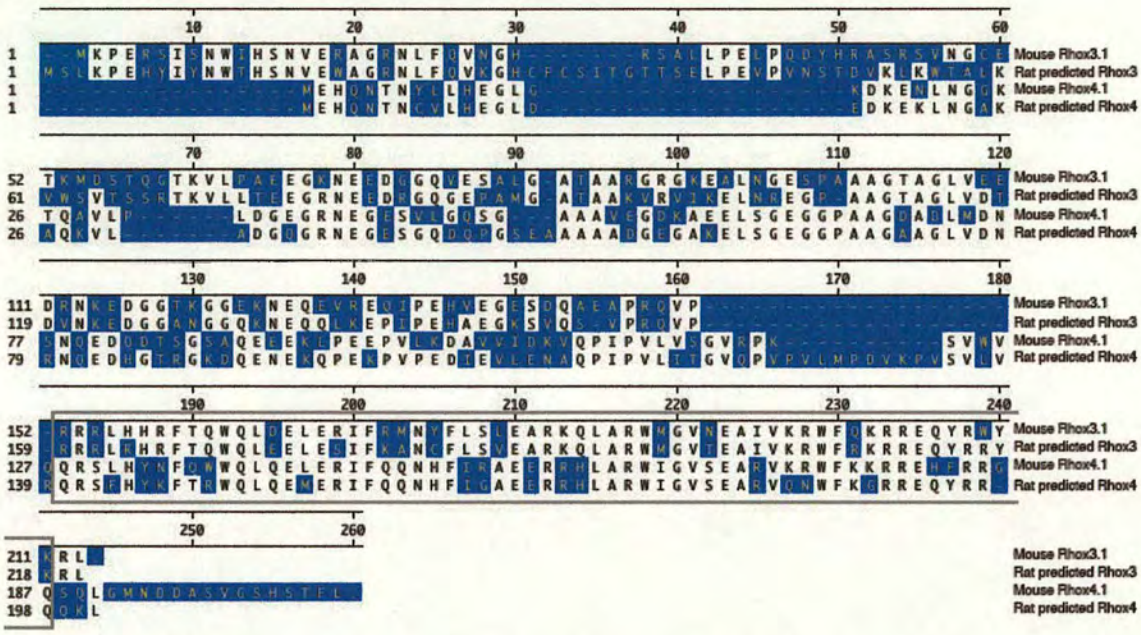


Figure 4.16 Multiple alignment of mouse and rat *Rhox3* and *Rhox4* predicted orthologues. Sequence differences are highlighted and the position of the homeodomain is contained within the box.

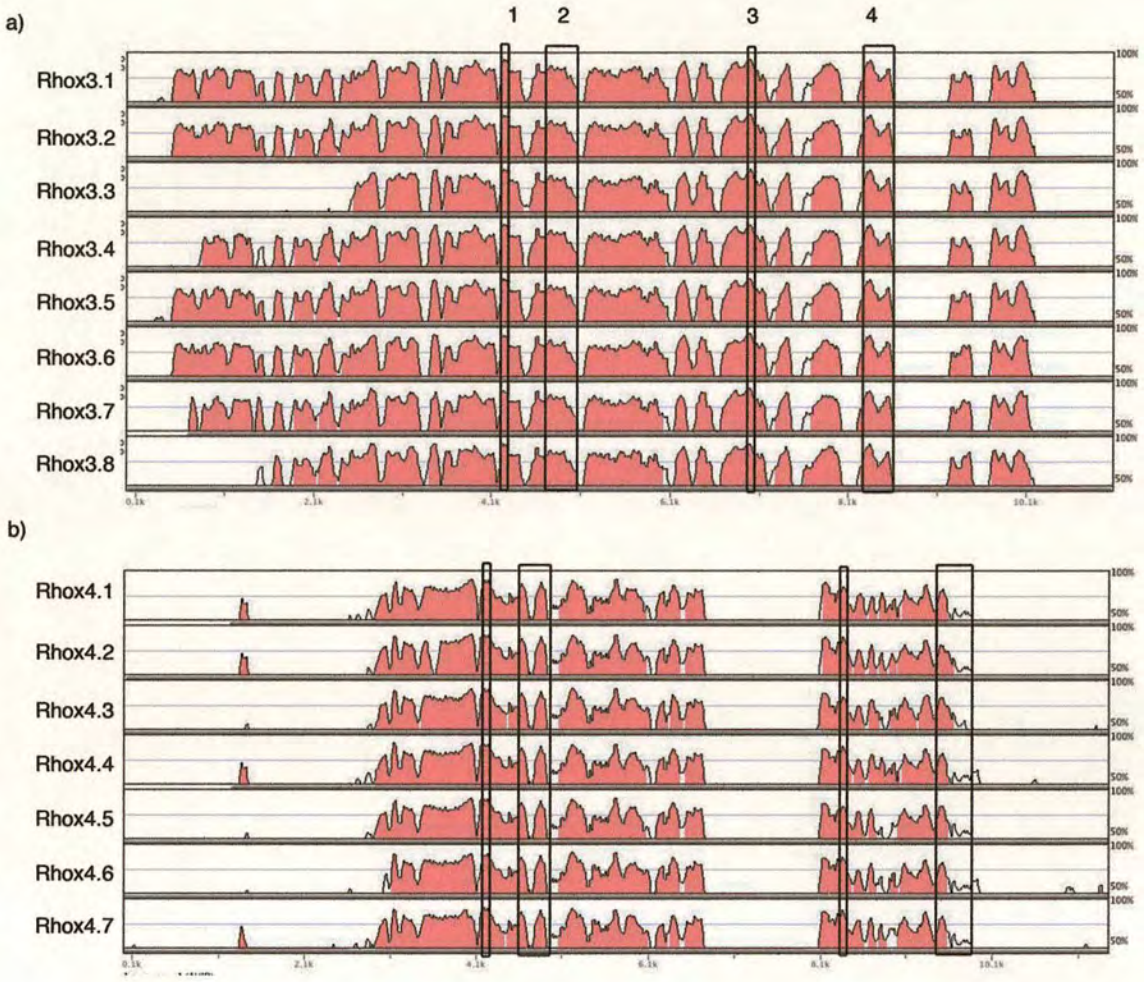


Figure 4.17 Rat and mouse *Rhox3* and *4* orthologues display some degree of sequence identity within intron, upstream and downstream sequences. AVID global sequence alignment of rat genomic sequence showing a) rat *Rhox3* with the eight mouse *Rhox3* gene copies and b) rat *Rhox4* with the seven mouse *Rhox4* gene copies. Percentage identity is shown on the vertical axis and position within the sequence on the horizontal axis. The position of the four exons in both genes are boxed and numbered.

orthologues were 67.1% identical over a region spanning from 4.2kb 5' of the start codon to 2.5kb 3' of the polyadenylation signal (Table 4.6d), and visualisation of the alignment showed that sequence similarities are present in intronic and upstream sequences and in a region of downstream sequence (Figure 4.17a). The *Rhox4* orthologue genomic sequences were less highly conserved at 52.8%, partially due tothe increased size of the rat *Rhox4* exon 2. AVID global alignment of *Rhox4* orthologous genomic sequences showed that sequence similarities were restricted to intron and upstream sequences; no significant sequence identity was found in downstream sequences (Figure 4.17b). Although rat and mouse *Rhox4* and *Rhox3* sequences are equally conserved at the peptide level, *Rhox4* mouse and rat genomic sequence has undergone more divergence than *Rhox3* orthologues. Finally in support of these orthologue predictions, mouse and rat *Rhox5* are also only 74% identical in peptide sequence (Sutton and Wilkinson 1997), indicating that the high level of sequence divergence present between mouse and rat orthologues of *Rhox3* and *Rhox4* is not unique.

To determine whether multiple copies of *Rhox4* are present in the rat, Southern analysis was carried out using a rat *Rhox4* specific probe (Table 4.7 and Figure 4.18). The presence of a single band of the predicted size in each digest strongly suggests that there is only one *Rhox4* copy in the rat. Additionally, to identify expression of any additional copies of rat *Rhox4,* shotgun cloning was carried out on RT-PCR products generated from a cDNA mix using two different sets of rat *Rhox4* primers at low stringency. Sequencing of 50 colonies failed to identify any sequence polymorphisms in *Rhox4*, again suggesting that there is only one copy of *Rhox4* in the rat. Since there are multiple copies of *Rhox4* in all sub-strains of mice investigated (Figure 4.7) these data suggest that the duplication event occurred either at the point of separation of the two species or before the divergence of the mouse sub-species investigated.

Two of the eight mouse *Rhox3* copies were predicted to produce a protein product lacking fifty three residues from the N-terminus. It was also unclear which start codon would be utilised by the remaining six copies. The predicted rat *Rhox3* is 220

residues, which is a similar size to the long mouse product of 213 residues. Alignment of mouse copy *Rhox3.1,* which is predicted to produce the longer product, with rat *Rhox3* (Figure 4.16) shows that the N-terminal region is conserved between the two species, furthermore, the second ATG at position 54 in the mouse sequence, suggested to be utilised by mouse *Rhox3.2* and *3.7* is lacking from the rat sequence. It is therefore likely that, if expressed, mouse *Rhox3* copies, *3.1, 3.3, 3.4, 3.5, 3.6* and *3.8,* lacking the stop codon downstream of the first ATG produce the longer product of 213 residues and that only copies *3.2* and *3.7,* that contain the stop codon, produce a truncated protein product.

Table 4.7 Sizes of fragments predicted after digestion with restriction endonucleases for predicted rat *Rhox4*.

Digest	EcoRI	EcoRV	NdeI	ScaI	SspI
Band size (kb)	8.0	7.3	2.9	3.4	1.8

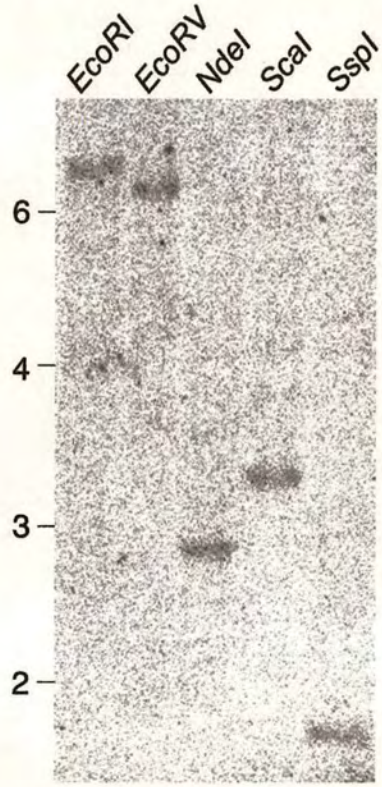


Figure 4.18 There is a single copy of *Rhox4* in rat. DA genomic DNA was digested with the restriction endonucleases shown and separated on a 0.8% agarose gel. After Southern blotting, a rat *Rhox4* exon 2 probe was used to reveal the restriction fragments. Bands confirmed the presence of a single gene copy, in agreement with predicted sizes. Position of size markers (kb) are indicated to the left of blot.

4.3 Discussion

In this chapter, I have described a duplication event which has resulted in a tandem array of three members of the *Rhox* family of homeobox-containing genes, *Rhox2, 3* and *4*. Eight complete copies of *Rhox2* and *3* are present and seven of *Rhox4*. All gene duplicates are conserved and are predicted to be functional. Rat orthologues of both *Rhox3* and *Rhox4* were identified, although the sequences were highly diverged from their mouse counterparts. Unlike in the mouse, *Rhox4* is present in a single copy in the rat genome.

Gene duplication mediated by unequal crossing over or transposable elements is widespread. However, due to relaxed selective pressure after duplication most gene duplicates are rapidly lost. The high degree of conservation of the *Rhox2, 3* and *4* copies is therefore highly surprising. Since there were multiple *Rhox4* copies in all sub-species of mice investigated, it is unlikely that the high degree of sequence identity can be explained by the duplication event having occurred very recently. In addition to the high conservation of coding regions, the intronic, upstream and downstream sequences were also conserved. In duplicates of all three genes, variable regions were limited to distinct regions within the intronic sequences and upstream and downstream of the genes. Although not highly divergent, these regions demonstrate that some sequence divergence has been allowed and had time to occur.

The tandem duplication of three of the four *Rhox* α cluster members results in a very gene rich region. The 24 genes of the *Rhox* α cluster are present within 350kb, which is an average of one gene per 14.2kb. In comparison, the mouse genome has on average one gene per 113.6kb (Waterston, Lindblad-Toh et al. 2002) and another reported gene rich region on mouse chromosome 13, containing three gene families has one gene per 53.6kb (Mallon, Wilming et al. 2004). Although the average region size containing one gene also reflects the gene size, and the genes of the *Rhox* α locus are small, each gene encompasses around 5kb, the genes are still tightly packed; there is an average of only 8.5kb between each of the genes within the duplication. The tight packing of genes may result in reduced gene expression at

these loci due to reduced DNA accessibility, interference between factors involved in gene regulation and competition for these factors (Chiaromonte, Miller et al. 2003). In turn these limitations may help maintain the DNA sequence identity of the duplicated copies as more than one functional gene copy may be required to produce enough protein product to fulfil the gene function.

Rhox2, 3 and *4* are present in the same duplicated unit, however, their copies appear to be diverging at different rates. The eight *Rhox2* copies have 93.8% peptide sequence identity. However, some *Rhox2* copies have diverged significantly more than others: *Rhox2.4* and *2.6* are only 87.5% identical in peptide sequence whilst *Rhox2.1* and *2.5* are 100%. The predicted cDNA sequences of *Rhox3* copies are highly conserved (98.3%), but two of the eight *Rhox3* copies are predicted to produce a protein product which is truncated by fifty three residues at the N-terminus, and reduces the average peptide sequence identity to 92.9%. As nothing is currently known about the function or interactions that take place with *Rhox* genes, it is unclear whether or not the truncated *Rhox3* product would affect the function of the protein despite the intact homeodomain. The level of sequence identity of *Rhox4* copies (mean of 97.3%) is much greater than *Rhox2* and *3*. Further work on the function and expression of the copies of *Rhox2, 3* and *4* is required to determine whether the differences in peptide and genomic sequence observed are resulting in the creation of pseudogenes or neo- and sub-functionalisation of gene function.

Gene conversion is widespread between gene duplicates (Murti, Bumbulis et al. 1992; Hogstrand and Bohme 1997; Noonan, Grimwood et al. 2004) and all three *Rhox* genes present in the duplicated unit contain evidence of gene conversion events. The peptide sequences of *Rhox2, 3* and *4* all contain polymorphisms which are present in multiple copies. However two thirds of the polymorphic residues of *Rhox4* are present in multiple copies, a much higher proportion than seen with the other two genes. Since *Rhox2, 3* and *4* are in the same duplicated unit the differences in copy diversity may reflect differing levels of selective pressure being applied to each of the three duplicated genes, suggesting this as an ideal locus for investigating evolutionary mechanisms. Additional work is required to determine whether similar

processes are occurring in the *Rhox* α cluster of sub-species of mice other than inbred C57BL/6 mice investigated in this study.

The presence of multiple copies of *Rhox4* in all mouse sub-species examined, but not in rat, suggests that there was either one duplication event in the mouse lineage, or a series of events that took place in quick succession at or soon after the time of divergence of the two species. Although there is currently little work on the role of the *Rhox* family, the *Rhox*es are all expressed in reproductive tissues, making them ideal candidates for driving a speciation event caused by reproductive isolation. The duplication event described in this chapter may therefore have played a role in the separation of the rat and mouse species. Rats and mice diverged between 12 and 25 million years ago (Adkins, Gelke et al. 2001; Springer, Murphy et al. 2003) and, whilst their genomes are clearly very similar in structure and content, they also contain many rearrangements and orthologous genes have undergone varying degrees of divergence (Waterston, Lindblad-Toh et al. 2002; Gibbs, Weinstock et al. 2004). On average mouse and rat genes are 94% identical in peptide sequence (Wolfe and Sharp 1993), making the finding that rat *Rhox3* and *Rhox4* display only 64% identity to their mouse orthologues surprising. The low level of sequence identity suggests that *Rhox3* and *4* have experienced periods of enhanced mutation rate, either due to relaxed selective pressure or positive selection for beneficial mutations. Since mouse *Rhox3* and *4* display a higher degree of sequence identity to their rat counterparts than they do to any of the other eleven mouse *Rhox* family members, it is unlikely that these orthologue predications are incorrect. Due to sexual selection and differences in reproductive mechanisms between species, many genes with roles in reproduction have undergone rapid divergence and positive selection (Sutton and Wilkinson 1997; Ting, Tsaur et al. 1998; Wyckoff, Wang et al. 2000). Examples include the *Rhox5* mouse and rat orthologues (Sutton and Wilkinson 1997) suggesting that this may be a common feature of *Rhox* family orthologues. Completion of the rat genome sequence is likely to uncover more *Rhox* genes, enabling further comparisons between rat and mouse family members.

Chapter 5 Expression of the *Rhox* α cluster

5.1 Introduction

Rhox family members are primarily expressed in reproductive and extra embryonic tissues, although three members are each expressed in one additional tissue, including *Rhox4* which is expressed in the developing and adult thymus (Jackson, Baird et al. 2002; Jackson, Baird et al. 2003; Maclean, Chen et al. 2005). Interestingly, *Rhox* genes are reported to be co-linearly expressed in the developing testis (Maclean, Chen et al. 2005). The discovery in Chapter 4 of a tandem duplicated unit in the *Rhox* α cluster (containing *Rhox2, 3* and *4*), raised the question of whether all copies are expressed.

In this chapter I describe the expression of the seven mouse *Rhox4* copies, and determine whether co-linear expression mechanisms are operating within the *Rhox* α locus during thymus organogenesis. Additionally, the expression of the rat orthologues of mouse *Rhox3* and *4,* and the two human *Rhox* family members are described.

5.2 Results

5.2.1 All Seven *Rhox4* copies are Expressed

As described in Chapter 4, the cDNA sequences of all seven *Rhox4* copies are highly conserved and all are predicted to be functional. To determine whether all seven copies of *Rhox4* are expressed, RT-PCR was carried out using a proofreading *Taq* polymerase and universal *Rhox4* primers on a mixture of RNA from all mouse tissues previously shown to express *Rhox4* (termed RNA mix) (Figure 5.1). The PCR product was shotgun cloned and sequenced, sequences were then aligned with those of the seven predicted *Rhox4* copies and the nucleotide present at each polymorphic base was noted. As predicted from the cDNA sequence analysis, sequences from all seven *Rhox4* copies were identified in the RNA mix (Figure 5.2a). However, copies were not present in equal quantities, *Rhox4.1, 4.2 and 4.7* made up 74% of transcripts seen and *Rhox4.3* was only detected once in seventy-six independent clones. This difference in expression levels could be due either to biased expression of individual copies in *Rhox4* expressing tissues or to expression of a single copy in individual tissues.

To establish whether different copies of *Rhox4* display different spatial patterns of expression, cDNA was isolated from four different mouse sources; 1) the pharyngeal region of E9.5 embryos (Figure 5.3); 2) adult thymus; 3) adult testis and; 4) E15.5 placenta. These individual tissues generally showed preferential expression from one or two copies (Figure 5.2b-e). The adult thymus and E9.5 pharyngeal region almost exclusively expressed *Rhox4.1* and *4.2,* although the ratio of expression of the two copies differed between these tissues, *Rhox4.1* expression was also dominant in the E15.5 placenta. The testis showed a more even expression of copies. However, uniquely amongst the tissues examined *Rhox4.5* was the most highly expressed in the testis, representing 33% of transcripts, suggesting that mechanisms controlling expression of *Rhox4* in the testis may be different from other tissues. *Rhox4.6* was only detected once in testis but multiple times in the cDNA mix, suggesting this copy is expressed primarily in sites other than those examined here.

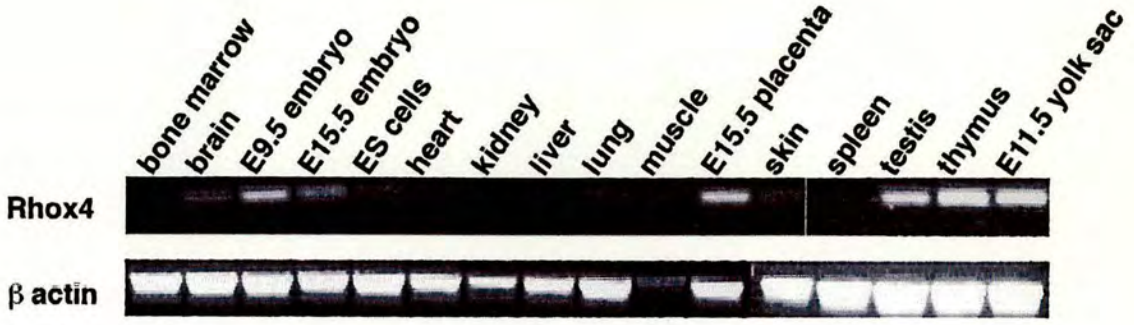


Figure 5.1 *Rhox4* expression can be detected in multiple embryonic and adult tissues. RT-PCR for *Rhox4* in embryonic and adult tissues at 35 cycles.

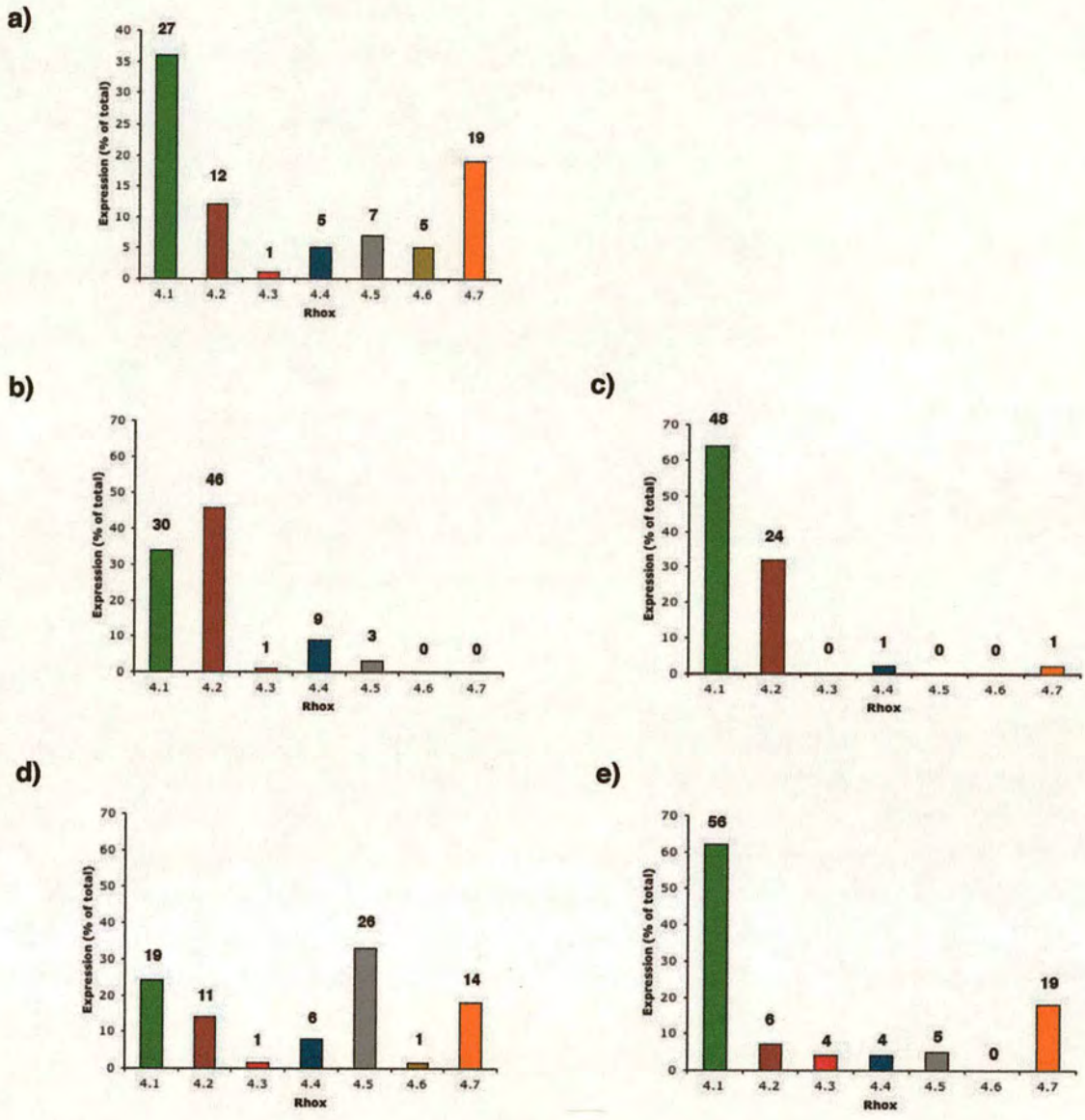


Figure 5.2 All *Rhox4* gene copies are expressed. Relative expression of each copy of *Rhox4* in a) cDNA made from a mixture of adult and embryonic tissues known to express *Rhox4* (76 clones in total), b) E9.5 pharyngeal region (89 clones), c) adult thymus (74 clones), d) testis (78 clones) and e) E15.5 placenta (91 clones). Expression was determined by shotgun cloning and sequencing of RT-PCR products generated after 32 cycles. Numbers indicate number of clones found to represent each *Rhox4* copy.

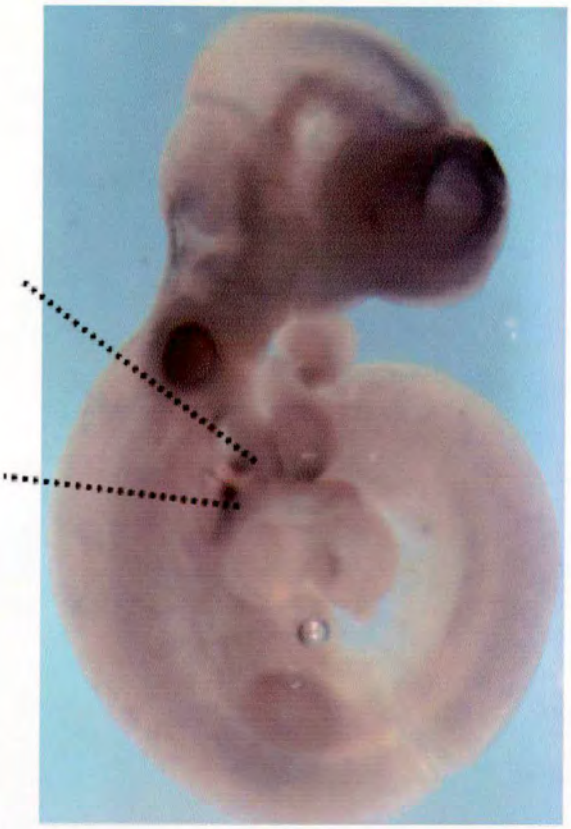


Figure 5.3 Image shows an E9.5 mouse embryo processed by *in situ* hybridisation using a full-length *Rhox4* probe. The lines indicate the region dissected for expression analysis (E9.5 pharyngeal region). The region contains the 2nd and 3rd pharyngeal pouches, surrounding mesenchyme, surface ectoderm and the neural tube.

This sequence analysis also revealed that 1% of *Rhox4.1*, 12% of *Rhox4.2 and 4.4* and 6% of *Rhox4.7* colonies contained sequences which lacked exon three. *Rhox4* exon three is only 48bp in length but encodes for sixteen residues of the homeodomain, including regions of the 2nd and 3rd helices and the loop region between them. Transcripts lacking this exon would therefore be predicted to be unable to bind DNA. However, the resulting transcript would be translated as the exon two to four splice is in frame. The cause of the omission of exon three in these four copies may be due to an inefficient splice acceptor site. However this has not been investigated further and it is unknown whether the product is biologically relevant.

5.2.2 Determination of *Rhox4* copy cDNA sequences

To determine the correct sequence of the untranslated regions and to identify whether they contained any polymorphic regions, 5' and 3' RACE were carried out on a mixture of RNA from tissues shown to express *Rhox4* by RT-PCR at 35 cycles (Figure 5.1). The 5' and 3' *Rhox4* specific primers used to amplify the final RACE products corresponded to regions in which all seven copies of *Rhox4* are identical.

5' RACE gave three different sized PCR products – a doublet of around 400bp and a single band around 200bp (Figure 5.4a). These were cloned and sequenced separately. The single, small band represented transcripts from *Rhox4.1, 4.4* and *4.5* (Figure 5.4b), and initiated five nucleotides upstream of the start codon (Figure 5.4d). Sequences from all copies except *Rhox4.4* were found in the double band (Figure 5.4c), and contained transcripts with untranslated regions up to 223bp in length (Figure 5.4d). The 5' UTR mainly consisted of a currently unique seven nucleotide repeat, TTTCGGC, differences in the number of repeats of this motif accounted for the differences in the length of the UTR between *Rhox4* copies. The length of the 5' UTR of each copy was not always identical and the UTRs shown in Figure 5.4d are the maximum identified for each copy. Since the numbers of colonies obtained for each copy were not large, it is not certain that the sequences shown in Figure 5.4d are the longest transcripts produced. Notably, the 5' UTR of

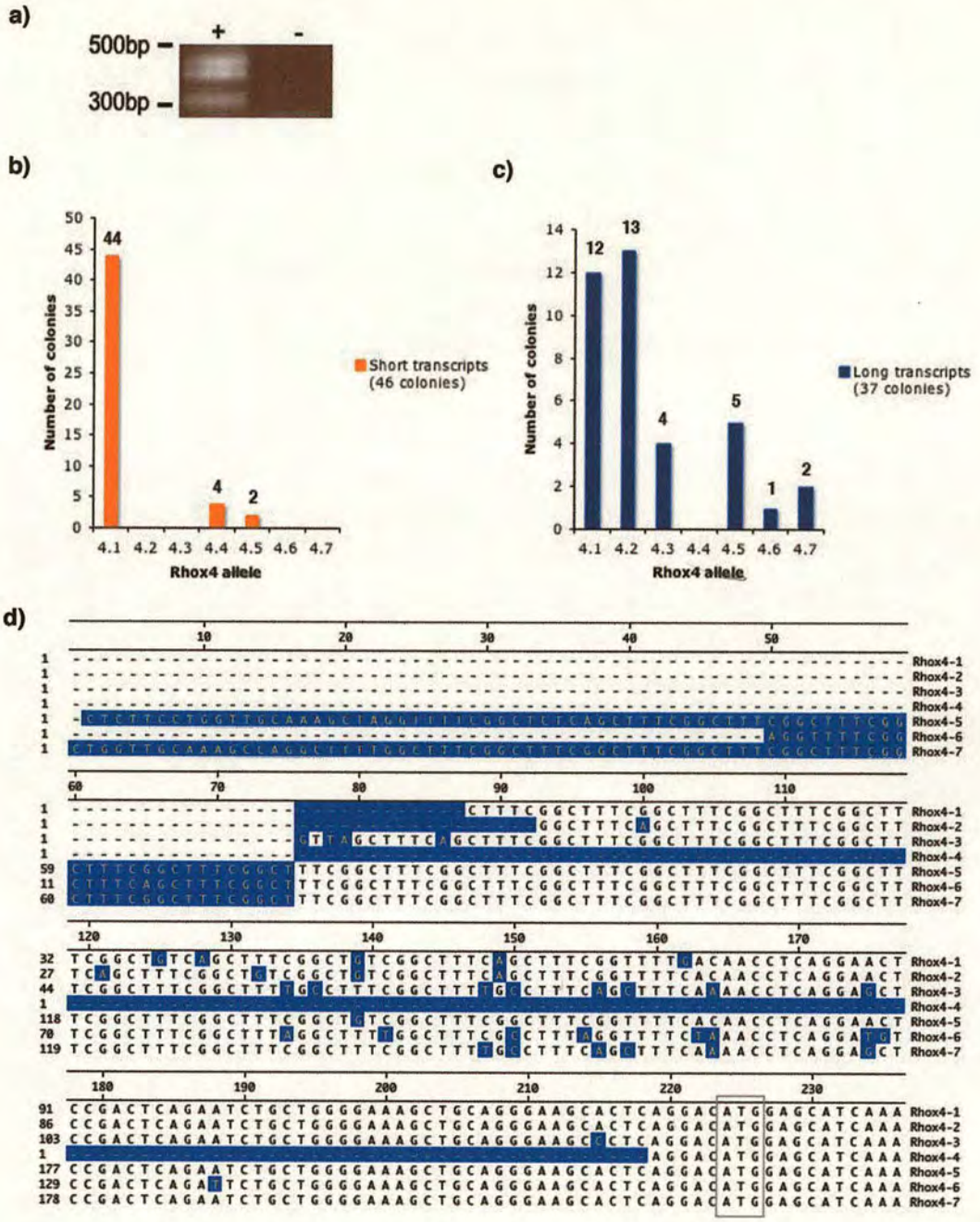


Figure 5.4 5' RACE using a universal *Rhox4* primer a) 5' RACE RT-PCR products generated with (+) and without (-) template cDNA. Position of size markers are indicated to the left of gel. b, c) Number of independent colonies that contained each *Rhox4* gene copy from the short (b) and long (c) transcripts. Number of colonies of each copy are indicated above bar. d) Multiple alignment of the longest 5' UTRs identified for each copy, the position of the start codon is contained within the box. Sequence differences are highlighted.

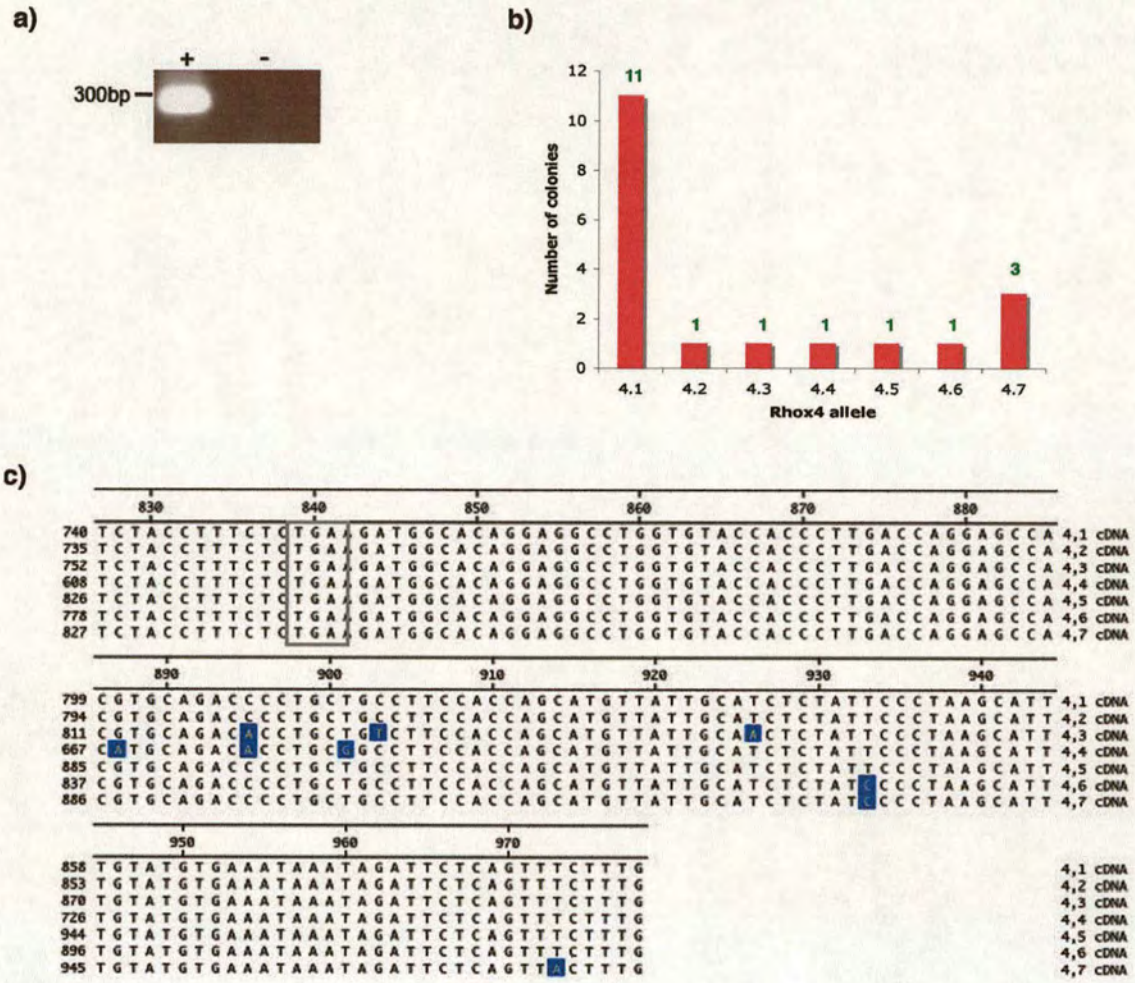


Figure 5.5 3' RACE using universal *Rhox4* primers a) 3' RACE RT-PCR products generated with (+) and without (-) template cDNA. Position of size marker is indicated to the left of gel. Number of independent colonies that contained a transcript from each *Rhox4* gene copy (number of colonies for each copy are indicated above bar). c) Multiple alignment of the 3' UTRs identified for each copy, the position of the stop codon is contained within the box. Sequence differences are highlighted.

Table 5.1 Accession numbers issued by EMBL of the seven *Rhox4* gene copy cDNA sequences determined by 5' and 3' RACE.

Rhox4 Gene Copy	Accession Number
Rhox4.1	AJ972665
Rhox4.2	AJ972666
Rhox4.3	AJ972667
Rhox4.4	AJ972668
Rhox4.5	AJ972669
Rhox4.6	AJ972670
Rhox4.7	AJ972671

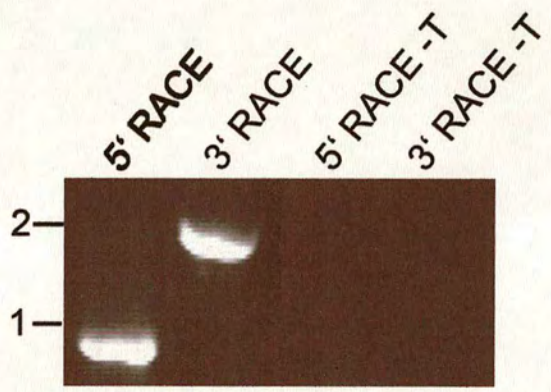


Figure 5.6 5' and 3' RACE products generated using β actin specific primers. No template controls labelled –T. Numbers to left of gel indicate size (kb). Expected product sizes are 872bp and 828bp (5' RACE) and 1800bp and 1715bp (3' RACE).

Rhox4.6 contained an upstream open reading frame of 114bp, which overlaps with the correct ORF and may therefore affect the expression of this copy.

A single product was seen in the 3' RACE (Figure 5.5a), the size of which agreed with published sequence (NM_021300) (Figure 5.5b,c). The cDNA sequences of the seven copies of *Rhox4* obtained from these 3' and 5' RACE analysis and were submitted to EMBL (Table 5.1).

The β actin control included with each 5' and 3' RACE reaction always gave a single band of the predicted size (Figure 5.6), indicating that the technique was working correctly and reliably.

5.2.3 The *Rhox* family does not display co-linear expression during thymus development

Each cluster (α, β and γ) of the *Rhox* family of homeobox-containing transcription factors has been reported to display quantitative co-linear expression during testis development. Additionally the α and γ clusters were reported to exhibit temporal co-linearity (Maclean, Chen et al. 2005). In agreement with published work (Jackson, Baird et al. 2003; Maclean, Chen et al. 2005), in this study only *Rhox4* expression was detected in the adult and developing thymus by qRT-PCR (Figure 5.7), although weak expression of *Rhox5* could be detect at 35 cycles by RT-PCR (Figure 5.8). Therefore, to determine whether the *Rhox* α cluster displays co-linear expression during thymus organogenesis, quantitative RT-PCR was employed using cDNA from stages of thymus development from E8.5 through E15.5 and adult. In these analyses, RNA was obtained from micro-dissected whole thymi; however, at E8.5, E9.5, E10.5 and E11.5 dissections all contained tissue surrounding the endoderm or thymus primordia and, as *Rhox4* is highly expressed only in the endoderm (Chapter 3), the level of expression in these samples appears artificially low. As expected, *Rhox4* was detected at all stages of thymus development (Figure 5.9). However no *Rhox1, 2* or *3*

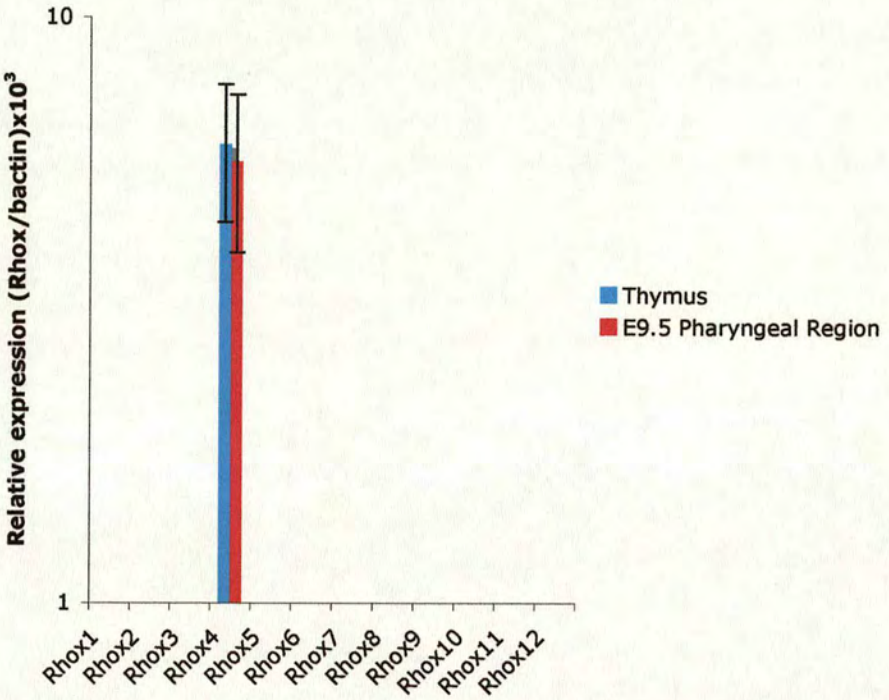


Figure 5.7 *Rhox4* is the only *Rhox* family member expressed in the developing and adult thymus. qRT-PCR of *Rhox* gene expression in the adult thymus and the E9.5 pharyngeal region which contains the recently formed 3rd pharyngeal pouch from which the thymus will form.

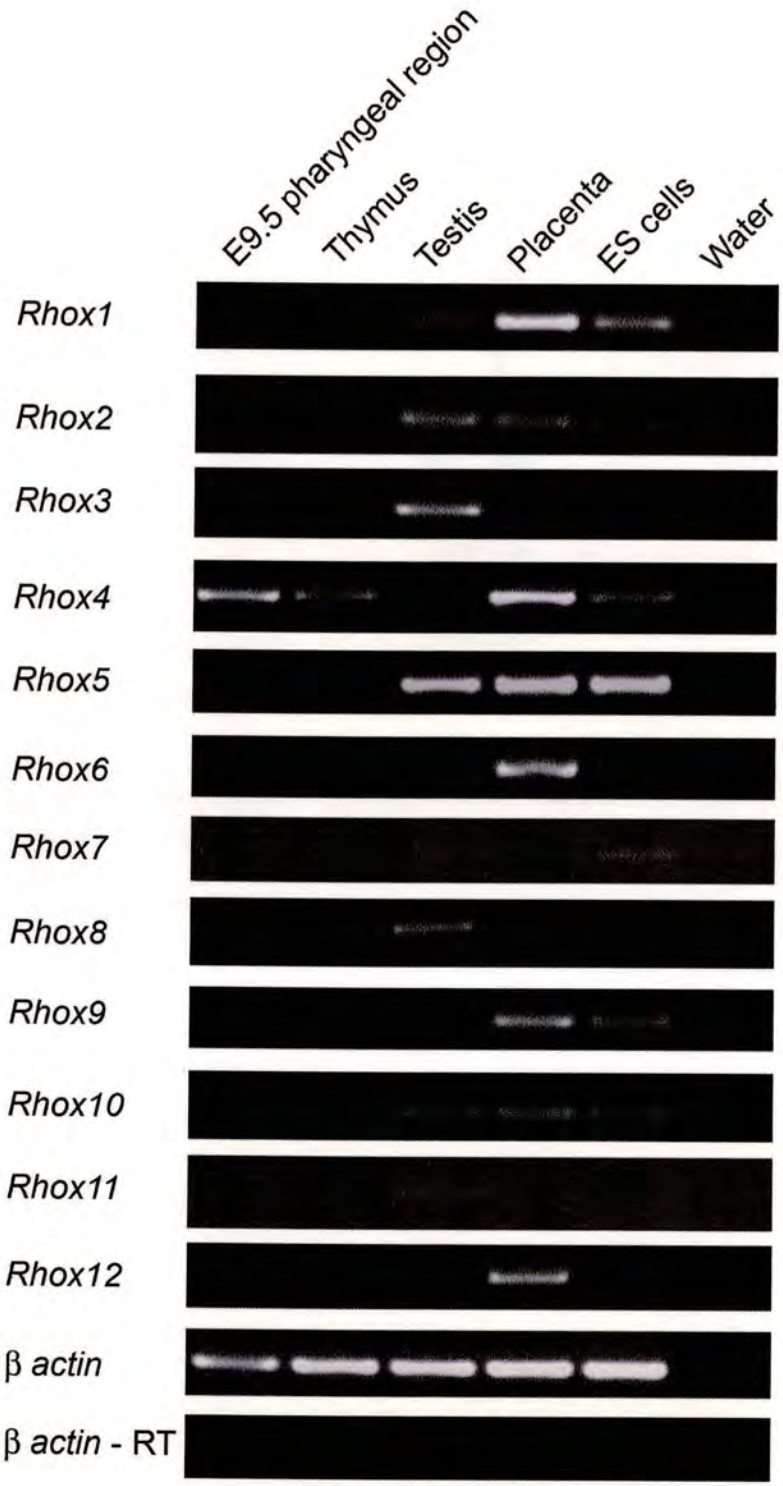


Figure 5.8 RT-PCR of *Rhox* gene expression showing that weak expression of *Rhox5,* in addition to *Rhox4* can be detected in the adult and developing thymus after 35 cycles.

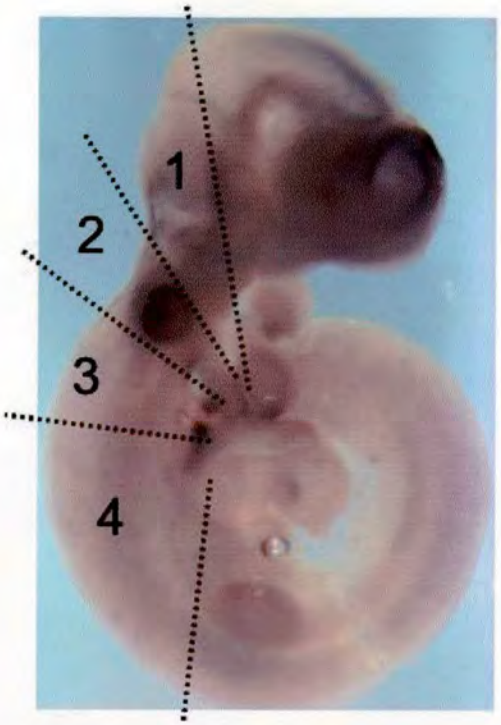
expression was found indicating that neither a temporal nor quantitative co-linear mechanism operates in these tissues.

To determine whether spatial co-linearity was displayed in the developing foregut at E9.5, the foregut was divided into four regions; Region 1 contained the most anterior region of the foregut endoderm, plus surrounding mesenchyme and ectoderm; Region 2 was as for Region 1 and also encompassed the region containing the 1st pharyngeal pouch; Region 3 contains the 2nd and 3rd pharyngeal pouches and surrounding tissues and; Region 4 contains the 4th pouch, foregut, midgut and hindgut (Figure 5.10). Again, only *Rhox4* expression was detected in the E9.5 gut tube; indicating the absence of spatial co-linearity in this tissue. These data indicate that co-linear expression mechanisms do not operate in the *Rhox* α cluster during thymus organogenesis.

5.2.4 Expression of rat *Rhox2* and *3*

In Chapter 3, I described rat orthologues of mouse *Rhox3* and *Rhox4*. To investigate whether the predicted rat *Rhox3* and *Rhox4* might also be functional orthologues, their expression pattern in different tissues was determined. Like their mouse counterparts both rat *Rhox3* and *Rhox4* were expressed in the placenta at E11.5 and in adult rat testis, although expression of *Rhox4* in the testis was weak (Figure 5.11). The thymus forms from the ventral portion of the 3rd pharyngeal pouch endoderm, which first begins to develop at around E9.5 of mouse embryogenesis. This region is marked by expression of *Rhox4* at E9.5 (Jackson, Baird et al. 2003) and, although strongly down regulated at E11, expression of *Rhox4* is always detectible in the embryonic and adult thymus by RT-PCR. However, rat *Rhox4* expression was not detected in either the pharyngeal region at E11.5 (the time at which the 3rd pharyngeal pouch forms in the rat), the E19.5 thymic primordia or the adult thymus. However, although mouse *Rhox4* is not expressed in the thymus at any stage of development, expression of rat *Rhox3* was found at all stages of thymus development investigated, similar to mouse *Rhox4*. *Rhox3* expression was also detected in the

a)



b)

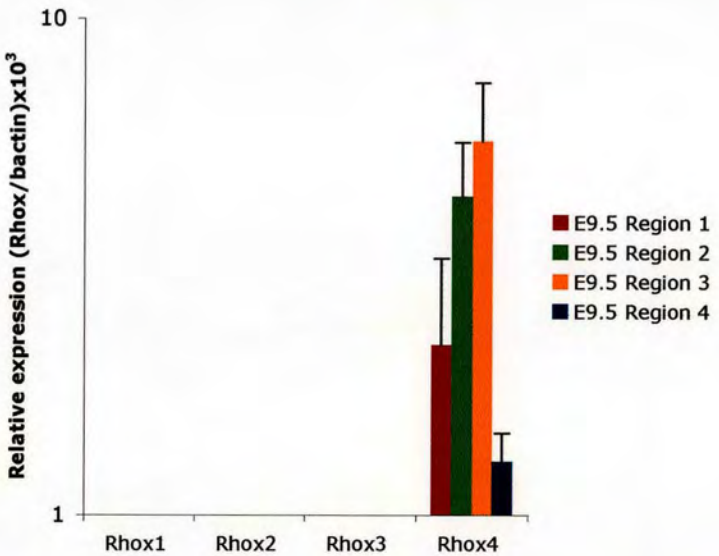


Figure 5.10 The *Rhox* α cluster does not exhibit spatial co-linear expression in the E9.5 gut tube. a) Image of an E9.5 embryo processed by *in situ* with a full length *Rhox4* probe showing the four regions used for b) qRT-PCR for *Rhox* α cluster genes in the E9.5 gut tube. Content of regions is described in the text, n=4.

Figure 5.11 Rat predicted *Rhox3* but not *Rhox4* is expressed in the developing and adult thymus. RT-PCR at 35 cycles on embryonic and adult rat tissues for predicted rat *Rhox3* and *Rhox4.* The E11.5 region is the pharyngeal region of rat embryos at the time when the 3rd pharyngeal pouch is formed.

brain, heart, muscle and skin, and mouse *Rhox4* expression can also be detected in these tissues by RT-PCR at 35 cycles (Figure 5.1). These data indicate that mouse *Rhox4* and rat *Rhox3* may be functional orthologues, and further suggest that there is a role for a *Rhox* family member in the thymus but that this is carried out by *Rhox4* in mouse and *Rhox3* in rat.

5.2.5 *Hoxa3, Pax1* and *Gcm2* are Expressed during Thymus Organogenesis in Human Embryos

Human foetal tissue used in the following work was obtained with approval from both the Lothian University Hospitals NHS Trust and the Lothian Research Ethics Committee.

There are two *Rhox* family members in humans, which are likely to share a single common ancestor with the mouse family of twelve *Rhox* genes (Maclean, Chen et al. 2005). Like the mouse *Rhox* family, human genes, *hPEPP1/GPBOX* and *hPEPP2* are expressed in the testis. However, in contrast to rat and mouse, no expression of either human *Rhox* family member was found in the embryonic thymic rudiment (Figure 5.12), and a previous publication failed to detect expression in the adult thymus (Wayne, MacLean et al. 2002). These analyses indicate that *Rhox* genes do not have a role in the human thymus after primordium formation.

We therefore carried out *in situ* hybridisation on first trimester human embryos to determine whether the expression patterns of transcription factors with established roles in the development of the mouse thymus are conserved in human embryogenesis. Like their mouse counterparts (see Chapter 3), expression of human *Hoxa3, Pax1* and *Foxn1* was detected in the pharyngeal region and/or thymic primordia of human embryos (Figure 5.13). *Hoxa3* expression was found throughout the 3rd pharyngeal pouch endoderm and surrounding mesenchyme at week six, the equivalent stage in mouse is E10.5. Further analysis is required to determine the anterior boundary of *Hoxa3* expression. Expression of human *Pax1* was found in all four pharyngeal pouches at week six, just after formation of the 4th pouch, suggesting

that the role of *Pax1* in the developing thymus is also conserved in human. Finally, *Foxn1* is expressed by the thymic primordia in human embryos at week seven, the equivalent of E11.5 to E12.5 in mouse embryogenesis. Further work is required to determine the time of onset of expression and whether it is expressed in a complementary domain with *Gcm2*.

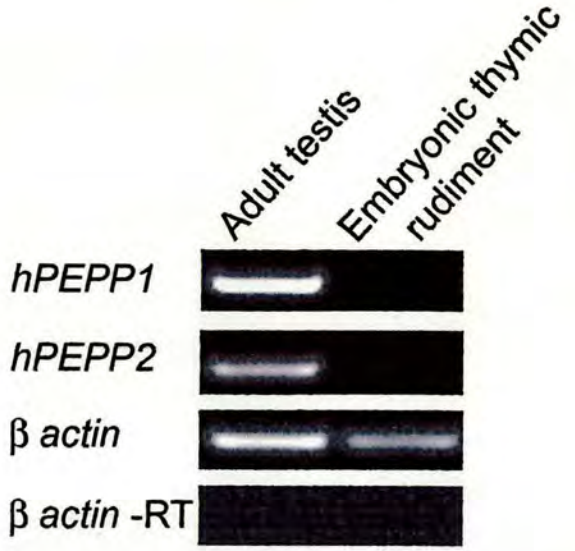


Figure 5.12 Human *Rhox4* family members, *hPEPP1* and *hPEPP2* are not expressed in the embryonic thymic rudiment. RT-PCR (35 cycles) for human *Rhox* family members and β actin on adult testis and a 7 week human embryonic thymus.

5.3 Discussion

In this chapter I have shown that all seven *Rhox4* copies are expressed, although some expression differences exist between tissues. The full-length cDNA sequences of each copy was determined and demonstrated that the 5' UTR of different copies contain variable copy numbers of a unique seven base repeat. A recent publication has suggested that *Rhox* gene expression in the developing testis is under the control of co-linear mechanisms. Due to the expression of *Rhox4* at all stages of thymus development, this analysis was extended to the expression of the *Rhox* α cluster genes during thymus organogenesis. However, no evidence was found to indicate that co-linear expression mechanisms were operating. The expression pattern of the rat orthologues of *Rhox3* and *Rhox4* were determined, and revealed that rat *Rhox3* but not rat *Rhox4* was expressed in all stages of rat thymus development investigated. Surprisingly, no expression of either human *Rhox* gene was found in the embryonic thymus. We therefore carried out initial expression profiling of transcription factors known to play a role in thymus organogenesis, to determine whether other differences in gene expression are present between mouse and human. However, *Hoxa3, Pax1* and *Foxn1* were all expressed in the pharyngeal region of human embryos in accordance to the mouse expression pattern.

Due to relaxed selective pressure after a duplication event most gene duplicates are rapidly lost. The very high degree of sequence identity of the seven *Rhox4* copies and eight *Rhox2* and *3* copies identified in Chapter 4 is therefore surprising. However, duplicates can be preserved if they gain a novel function (neo-functionalisation) or the expression and/or functions of the original gene are divided between the duplicates (sub-functionalisation). The data on *Rhox4* copy expression in different tissues illustrates that there may be functional differences between the seven copies. These are likely to be the result of on-going evolutionary processes, which may eventually result in the loss of some copies from the genome. The different expression levels of *Rhox4* copies in different tissues may reflect ongoing sub-functionalisation. *Rhox4.1* and *4.2* were highly expressed in the region encompassing the 3rd pharyngeal pouch and the adult thymus, whereas *Rhox5* was

mainly expressed in the testis. Identification of elements necessary for expression of *Rhox4* in these tissues will give an insight into the ability of individual copies to be expressed in each tissue.

The untranslated regions of mRNAs play important roles in control of translation by modulating mRNA stability and interaction with factors controlling translation and subcellular location (Mignone, Gissi et al. 2002). Due to their ability to activate a large number of genes, translation of transcription factor mRNAs is often tightly regulated by means of the UTRs, and the 5' UTR is often longer than average (Kozak 1987). With the exception of *Rhox4.4, Rhox4* copies were found to have maximum length 5' UTRs of between 131 and 223bp, which is around the average of 186.3bp established for rodent genes (Mignone, Gissi et al. 2002). In common with many other genes, the 5' UTRs were found to contain a repeated element, although the repeat appears to be unique to this gene. The 5' UTR of *Rhox4.6* contained an upstream ORF which may be the cause of the low expression of this copy. Transcripts from *Rhox4.1, 4.3* and *4.4* were also found which initiated only five nucleotides upstream of the start codon. Although we have not validated this result by a different method, it is unlikely that it is an artefact since many independent colonies were identified containing this exact sized transcript and 5' RACE for a control gene carried out simultaneously yielded bands of the predicted sizes. The minimum reported rodent 5' UTR length is 16 nucleotides (Mignone, Gissi et al. 2002) and, although it has been possible to initiate transcription with only a single nucleotide upstream of the AUG in a mammalian cell free system (Hughes and Andrews 1997), it is unclear whether this is sufficient *in vivo*.

Homeobox-containing genes have a conserved 60 amino acid motif that can bind DNA and regulate transcription. *Rhox4* belongs to a family of homeobox genes that are present in three adjacent clusters on the mouse X chromosome. The presence of the *Rhox* family as a cluster indicates that the family may have arisen by gene duplication and divergence from a common ancestor (Maclean, Chen et al. 2005). The preservation of families of homeobox genes in clusters rather than spread throughout the genome suggests that they may share common regulatory elements

and, in line with this, expression of the *Rhox* family in the developing testis has been shown to be regulated by both quantitative and temporal co-linear mechanisms (Maclean, Chen et al. 2005). Co-linear expression regulation in the *HoxD* cluster is controlled by a remote enhancer (Kmita, Fraudeau et al. 2002). A similar mechanism has been hypothesised to exist upstream of each of the three *Rhox* gene clusters (Maclean, Chen et al. 2005). However it is hard to imagine how this mechanism may function when the 2nd, 3rd and 4th genes in the α cluster are present in a tandem array, and all copies of the 4th gene, *Rhox4* are expressed. It is possible that *Rhox4.5,* which was the most highly expressed *Rhox4* copy in the testis is present at the optimal distance from a putative testis-specific enhancer, although this would not fully explain the pattern of copy expression in the testis. The study presented here establishes that co-linear expression mechanisms do not operate within the *Rhox α* cluster during thymus organogenesis and indicates that expression of *Rhox4* in the thymus is likely to be controlled by different elements than those responsible for expression in the testis.

The finding that *Rhox3* but not *Rhox4* is expressed in the developing and adult thymus in the rat is surprising. Further work is required to determine whether expression of rat *Rhox3* is found solely in the ventral foregut endoderm and progressively restricted to the thymic domain of the 3rd pharyngeal pouch, as seen for *Rhox4* in the mouse. It is also interesting that rat *Rhox3* expression can be detected in the brain, heart, muscle and skin, as low level expression of mouse *Rhox4* can also be detected in these tissues by qRT-PCR (Jackson, Baird et al. 2002) and RT-PCR (this chapter). These data suggest that mouse *Rhox4* and rat *Rhox3* may be functional orthologues, rather than mouse and rat *Rhox4* as clearly predicted by sequence identity. The difference in *Rhox* orthologue gene expression between the rat and mouse species may have been caused by the tandem duplication present within the mouse locus. However, how this may have occurred is unclear as the position of regulatory elements is not currently known. This result also suggests that *Rhox* family members have common downstream targets, although none have yet been identified.

The expression of one *Rhox* gene in the thymus of both mouse and rats suggests that they may have a conserved role in this tissue, however neither of the two human *Rhox* genes are expressed in the embryonic (this chapter) or postnatal thymus (Wayne, MacLean et al. 2002). Thus, either the hypothesised common ancestral gene (Maclean, Chen et al. 2005) was not expressed in the thymus and expression has been picked up by a rodent ancestor, or expression has been lost by the two human genes, PEPP1/OTEX (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002) and PEPP2 (Wayne, MacLean et al. 2002). This result suggests that there is a murine-specific role for an *Rhox* family member during thymus organogenesis. Finally, we show that *Hoxa3, Pax1* and *Foxn1* are all expressed during thymus organogenesis in human. Their expression patterns in the stages analysed appeared to correlate with expression of their mouse orthologues, suggesting that genetic mechanisms governing thymus development are conserved between mouse and human. No differences in the morphology of the 3rd pharyngeal pouch are present between human and mouse (A. Farley, unpublished) (Weller 1933; Norris 1938). However, after separation from the pharynx the human thymic primordia are enlongated structures which extend caudally, and consist of many lobules enclosed in a sheet of mesenchymal tissue (Norris 1938). In mouse the thymic primoria migrate caudally as a ball of tissue with no obvious lobular structure. However, the significance of these morphological differences is unclear. *Rhox4* may have a role in one or more of these differences in the thymic primordia of mouse and human, as *Rhox4* expression is found after primordia formation. However, due to expression of *Rhox4* in the mouse but not human 3rd pharyngeal pouch at E9.5, differences in formation of the common thymus/parathyroid primordia may also exist.

Chapter 6 Development of an Experimental Model for Investigation of the Role of *Rhox4* in Thymus Organogenesis

6.1 Introduction

Rhox4 is currently the only gene known to specifically mark the thymic domain of the 3rd pharyngeal pouch before the onset of *Foxn1* expression at E11.25. Three lines of evidence suggest that *Foxn1* is unlikely to specify the thymus: i) *FoxnI* null mice contain a developmentally arrested thymic rudiment that expresses thymic epithelial progenitor cell markers (Blackburn, Augustine et al. 1996; Bennett, Farley et al. 2002; Gill, Malin et al. 2002), ii) grafting of E8.5 – E9.0 2nd and 3rd pharyngeal pouch endoderm under the kidney capsule of nude mice gives rise to a functional thymus, showing that the endoderm is specified before onset of *Foxn1* expression (Gordon, Wilson et al. 2004) and iii) expression of IL-7, a TEC specific marker which is required for T cell development, is independent of *Foxn1* and can be detected in *Foxn1*$^{-/-}$ embryos at E12 (Zamisch, Moore-Scott et al. 2005). The progressive restriction and abrupt down-regulation of *Rhox4* suggests an early role in thymus organogenesis and would be consistent with lineage specification.

The complex organisation of the *Rhox4* locus described in chapters 4 and 5, means that this locus is not easily amenable to analysis via conventional genetic manipulation approaches. The discovery of RNA interference (RNAi) pathways in *C. elegans* (Fire, Xu et al. 1998) and subsequently mammals (Elbashir, Harborth et al. 2001) has revolutionised the use of antisense technology for gene-specific knockdown. There are now many reports of successful knockdown of gene expression in mice, using both exogenous small-interfering RNAs (siRNA) and by transgenic expression of short hairpin RNAs (shRNA) (Dykxhoorn and Lieberman 2005). RNAi is triggered by the presence of double-stranded RNA (dsRNA), which is cleaved into siRNA molecules of around 21 nucleotides in length by an RNase III

family member, Dicer (Bernstein, Caudy et al. 2001). Gene silencing is triggered by interaction of the siRNA with the RNA-induced silencing complex (RISC) (Hammond, Bernstein et al. 2000; Zamore, Tuschl et al. 2000), which is thought to result in the destruction of the complementary mRNA and inhibition of translation. RNAi expression systems have been designed that mimic endogenous micro-RNA (miRNA) silencing mechanisms by expressing a 70bp shRNA consisting of a 19-22bp sequence from the gene of interest followed by a short loop region and the corresponding antisense. RNA polymerase III promoters, human U6 or HI are most commonly used for shRNA expression (Dykxhoorn and Lieberman 2005). This approach is suitable for investigating the function of *Rhox4*, as shRNAs can be designed that should knock down the expression of all seven copies.

In this chapter, I describe the construction of two targeting cassettes designed to express *Rhox4* specific shRNA from the ROSA26 locus (Zambrowicz, Imamoto et al. 1997; Soriano 1999). One cassette mediates ubiquitous expression, and a second cassette should permit expression of the shRNA only after *Cre*-mediated recombination, enabling spatially and temporally controlled knockdown of *Rhox4*.

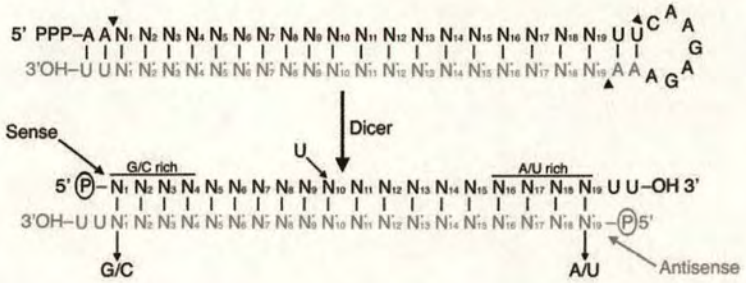
6.2 Results

6.2.1 Design of *Rhox4* specific RNAi sequences

In order to investigate the function of *Rhox4* in thymus organogenesis, shRNAs were designed, as described (Kunath, Gish et al. 2003; Sharpe and Mason 2005), to target *Rhox4* specific sequences. Briefly, 19bp *Rhox4* specific sequences were selected such that the G/C content of the 5' end of the sense sequence was higher than the 3' end. This is to maximise incorporation of the antisense siRNA strand into the RISC complex, since RISC incorporates Dicer products from the end with the lowest thermodynamic stability (Khvorova, Reynolds et al. 2003; Schwarz, Hutvagner et al. 2003) (Figure 6.1a). Sequences were chosen in regions displaying 100% identity between the seven *Rhox4* copies to ensure knockdown of all *Rhox4* copies. Three sequences predicted to mediate efficient knockdown were chosen (*Rhox4* RNAi1, 2 and 4, Figure 6.1b) and, in addition one ineffective sequence (*Rhox4* RNAi5) was chosen to provide a control for specificity of knockdown phenotypes.

To enable the knockdown efficiency of the chosen sequences to be tested, oligonucleotides containing the four shRNA sequences were annealed and cloned into a vector containing the H1 promoter as described (Kunath, Gish et al. 2003; Sharpe and Mason 2005) (Figure 6.1c). Quantification of knockdown efficiency was attempted by transient transfection of shRNA-expressing constructs into an ES cell line reported to express *Rhox4* (Jackson, Baird et al. 2002), followed by real-time RT-PCR analysis for *Rhox4* expression. The experiment was carried out by transient transfection of cells with each shRNA-expressing construct and GFP. After forty-eight or seventy-two hours cells were harvested and *Rhox4* expression was determined by real-time RT-PCR (Table 6.1). Although the results indicated that RNAi1 and 2 might mediate knockdown of *Rhox4* expression, at seventy-two hours the control RNAi5 also appeared to knockdown *Rhox4* expression. Whilst it is possible that RNAi number 5 does knockdown *Rhox4* expression, this result is more likely to be due to the level of *Rhox4* expressed by the ES cell line being too low to provide sufficient sensitivity to the assay. Thus, due to the extremely low levels of

a)



b)

		Asp718	SENSE	LOOP	ANTISENSE	TEMINATION	XbaI
Rhox4 RNAi1	5′	GTACC	AACACCAACTACCTACTTCAT	TTCAAGAGAAA	ATGAAGTAGGTAGTTGGTGTT	TTTTTGGAAAT	3′
	3′	G	TTGTGGTTGATGGATGAAGTA	AAGTTCTCTTT	TACTTCATCCATCAACCACAA	AAAAACCTTTAGATC	5′
Rhox4 RNAi2	5′	GTACC	AAGCTGCCAGAGGAGCCAGTT	TTCAAGAGAAA	AACTGGCTCCTCTGGCAGCTT	TTTTTGGAAAT	3′
	3′	G	TTCGACGGTCTCCTCGGTCAA	AAGTTCTCTTT	TTGACCGAGGAGACCGTCGAA	AAAAACCTTTAGATC	5′
Rhox4 RNAi4	5′	GTACC	GGGACAAAGCAGAAGAATTAA	TTCAAGAGAAA	TTAATTCTTCTGCTTTGTCCC	TTTTTGGAAAT	3′
	3′	G	CCCTGTTTCGTCTTCTTAATT	AAGTTCTCTTT	AATTAAGAAGACGAAACAGGG	AAAAACCTTTAGATC	5′
Rhox4 RNAi5	5′	GTACC	TTTAAGAAGAGGAGAGAGCAC	TTCAAGAGAAA	GTGCTCTCTCCTCTTCTTAAA	TTTTTGGAAAT	3′
	3′	G	AAATTCTTCTCCTCTCTCGTG	AAGTTCTCTTT	CACGAGAGAGGAGAAGAATTT	AAAAACCTTTAGATC	5′

c)

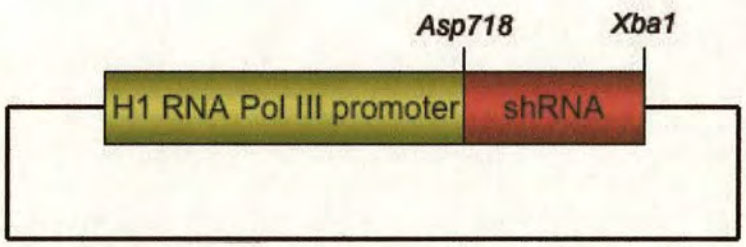


Figure 6.1 *Rhox4* RNAi design a) diagram of a shRNA and a siRNA molecule. Arrowheads in shRNA indicate possible Dicer cleavage sites and nucleotide preferences for effective RNAi are shown on the siRNA, diagram taken from (Sharpe and Mason 2005). b) Sequences of oligonucleotides chosen to mediate knockdown of *Rhox4*. The restriction sites, sense, loop, antisense and termination sequences are indicated. c) shRNA expression construct containing the human H1 RNA Polymerase III promoter. shRNA sequences shown in b) were inserted via the *Asp718* and *Xbal* restriction endonuclease sites.

Table 6.1 Efficiency of *Rhox4* RNAi expression constructs. E14tg2a ES cells were transiently transfected with both an *Rhox4* RNAi expression construct and GFP over-expression construct. Cells were incubated for 48 or 72 hours, harvested and sorted into GFP positive (+) and negative (-) fractions. *Rhox4* expression was determined by real-time RT-PCR. Rhox4 and β actin values are an average of one experiment where each reaction was carried out in duplicate.

Sample	Rhox4		β actin		Relative expression (Rhox4/βactin)x10^3		Relative expression (+/-)
	+	-	+	-	+	-	
48h RNAi1	443	407	136050	44955	3.2	9.0	0.3
48h RNAi2	204	227	40650	21060	5.0	10.0	0.5
48h RNAi4	220	626	38560	74935	5.7	8.0	0.7
48h RNAi5	492	102	64885	12495	7.5	8.1	0.9
72h RNAi1	415	2707	157050	128550	2.6	21.0	0.1
72h RNAi2	1558	1008	485900	156350	3.2	6.4	0.5
72h RNAi4	2550	1109	503850	207350	5.0	5.3	0.9
72h RNAi5	1183	303	213650	16675	5.5	18.2	0.3

Rhox4 expressed by the ES cell line, knockdown of *Rhox4* expression by the shRNAs could not be demonstrated.

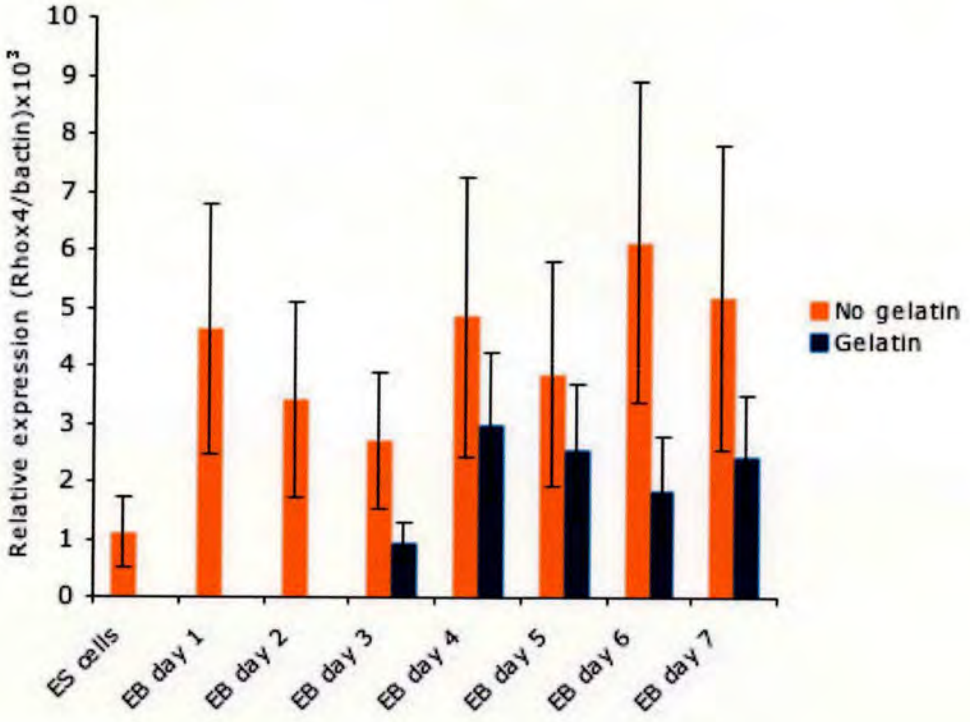
Rhox4 expression is reported to be upregulated 5-fold after seven days of *in vitro* differentiation of embryoid bodies (Jackson, Baird et al. 2002), suggesting that quantification of *Rhox4* expression in stable shRNA-expressing cell lines, after ES cell differentiation may provide a suitable validation strategy. To assess this possibility, we determined the levels of *Rhox4* expression in embryoid bodies generated using two differentiation protocols at different time points. In keeping with previous a report (Jackson, Baird et al. 2002), up-regulation of *Rhox4* was observed, although this occurred earlier in the culture than reported (Table 6.2, Figure 6.2). The level of *Rhox4* expression during differentiation was higher in suspension cultures than adherent cultures. However, the results obtained were highly variable by real-time RT-PCR at each time point, due to the very low real levels of *Rhox4* expression in the cultures. Thus, although *Rhox4* is up-regulated during ES cell differentiation, the real level of up-regulation is very small. This suggests that *in vitro* differentiation of ES cells by embryoid body formation will not be an effective method for screening ES colonies for *Rhox4* knockdown.

As no cell line expressing high levels of *Rhox4* is currently available an *Rhox4* over-expression cassette was constructed for this purpose. The over-expression cassette was constructed by RT-PCR amplification of an *Rhox4* product that included the start and stop codons. This was cloned into the TOPO2.1 vector, the insert was then excised and cloned downstream of the strong, ubiquitous promoter, CAGGS (Sharp, Kost et al. 1989). The orientation of the cloned fragment relative to the promoter was determined by sequencing, and expression of *Rhox4* was demonstrated by transient transfection in COS cells (Figure 6.3). Future work will use the *Rhox4* over-expression construct to make a stable cell line. Once constructed, the *Rhox* over-expressing cell line can be used to determine the knockdown efficiency of the *Rhox4* RNAi sequences via transient transfection with the shRNA expression constructs.

Table 6.2 *Rhox4* is up-regulated during *in vitro* differentiation of embryoid bodies. E14tg2a ES cells were allowed to form embryoid bodies by handing drop culture (Days 1 and 2); they were then transferred to either a suspension (- gelatin) or adherent culture (+ gelatin) and harvested at daily intervals and assessed for *Rhox4* expression by real-time RT-PCR. Rhox4 and β actin values shown are an average reading of 3 independent real-time RT-PCR experiments; samples in each experiment were carried out in duplicate.

Sample	Rhox4	β actin	Relative expression ratio (Rhox4/βactin)x10^3	Standard deviation of ratio
ES cells	313	283516	1.1	0.6
EB day 1	77	16991	4.6	2.1
EB day 2	50	14614	3.4	1.6
Day 3 - gelatin	84	30670	2.7	1.1
Day 4 - gelatin	504	105937	4.8	2.4
Day 5 - gelatin	302	79802	3.8	1.9
Day 6 - gelatin	402	65518	6.1	2.7
Day 7 - gelatin	288	55713	5.2	2.6
Day 3 + gelatin	48	47921	1.0	0.3
Day 4 + gelatin	313	104841	3.0	1.2
Day 5 + gelatin	304	121490	2.5	1.1
Day 6 + gelatin	240	133223	1.8	0.9
Day 7 + gelatin	327	136065	2.4	1.0

a)



b)

Figure 6.2 Rhox4 is up-regulated during in vitro differentiation of embryoid bodies (EB). E14tg2a ES cells were allowed to form embryoid bodies by hanging drop culture (Days 1 and 2); they were then transferred to either a suspension (no gelatin) or adherent culture (gelatin), harvested at daily intervals and assessed for *Rhox4* expression by a) real-time RT-PCR (n=3, one experiment) or b) RT-PCR (non-quantitative).

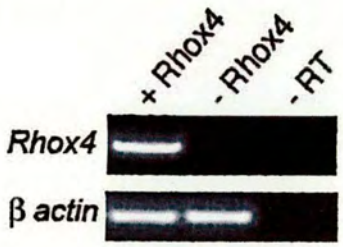


Figure 6.3 Expression of *Rhox4* from the CAGGS promoter in transiently transfected COS cells. *Rhox4* was inserted using a single restriction site, colonies were therefore obtained that contained *Rhox4* in both the correct (+ Rhox4) and incorrect (-Rhox4) orientations.

6.2.2 Generation of a shRNA expression construct for *Rhox4* knockdown

The shRNA expression construct for ubiquitous *Rhox4* knockdown contained the *Rhox4* shRNAs described above downstream of the human HI promoter. A *flpe*-recombinase removable selection cassette (neomycin and thymidine kinase) was included to allow selection of targeted ES cell clones and subsequent removal of the cassette. Additionally, a cassette containing the GFP cDNA under control of the CAGGS promoter was included, to allow shRNA-expressing clones to be visualised. Strong expression of GFP in all shRNA-expressing cells should enable chimeric animals to be analysed for contribution of knockdown cells to each tissue. The CAGGS-GFP was flanked by *loxP* sites to allow it to be excised. To ensure ubiquitous expression of the shRNAs and to minimise position effects, the cassette was cloned into a ROSA26 targeting construct (Zambrowicz, Imamoto et al. 1997; Soriano 1999) (Figure 6.4). Targeting of the construct into a known locus rather than introducion by random transgenesis will permit direct comparison between *Rhox4* specific and control shRNAs. Additionally, ensuring that the expression construct is present in a single copy allows removal of the selection cassette by FLPe mediated recombination, without the possibility of rearrangements occurring. Rearrangements are possible in ES cell lines generated by additive transgenesis since, multiple, tandem arrayed transgene insertions are common. The construct was made via the following steps (Figure 6.4 See Appendix for plasmid maps); steps 4 and 5 have yet to be completed:

1) *Introduction of the human HI promoter into pBluescript.* This was carried out by a) ligation of annealed oligonucleotides containing an *EcoR1* sticky end, a *PacI* site, the 100bp human HI promoter, an *Asp718* site and an *XbaI* sticky end into the *EcoRI* and *XbaI* site of pBluescript and b) introducing a second olignucleotide containing the restriction sites *BamHI* and *AscI* into the *XbaI* site and the *SacII* site of pBluescript.

2) *Ligation of the shRNA containing oligonucleotides downstream of the HI promoter.* The four *Rhox4* shRNA containing oligonucleotides shown in Figure 6.2 were ligated into the product of step 1 via the *Asp718* and *XbaI* sites in both insert and vector, generating one construct for each shRNA sequence. All further steps were carried out in parallel with all four constructs.

3) *Construction and introduction of loxP-CAGGS-GFP-loxP.* A cassette containing CAGGS-GFP was isolated from pPyCAGGFPIP (a gift from Prof Austin Smith, ISCR) by *EcoRV* digestion, and was inserted between the two LoxP sites of pROSA by blunt end cloning (Figure 6.5a). A further cassette containing a polyA signal and a transcriptional pause (Ashfield, Patel et al. 1994) were then inserted 3' of the GFP also by blunt end cloning (Figure 6.5b). The *loxP-CAGGS-GFP-loxP* was then ligated into the product of step two by digestion of the backbone with *XbaI* and the insert with *AvrII*.

4) *Insertion of the selection cassette.* A *frt*-neomycin-thymidine kinase-*frt* cassette was obtained from Dr Andrew Smith (ISCR), the *NotI* site was removed by digestion with *NotI* and religation after blunt end cloning. The resulting cassette will be ligated into the product of step three by the *BamHI* site in both backbone and insert.

5) The shRNA expression construct will be inserted between the ROSA26 targeting arms (Zambrowicz, Imamoto et al. 1997; Soriano 1999) via the restriction sites, *PacI* and *AscI*. The construct can be linearised via *NotI* for targeting ES cells.

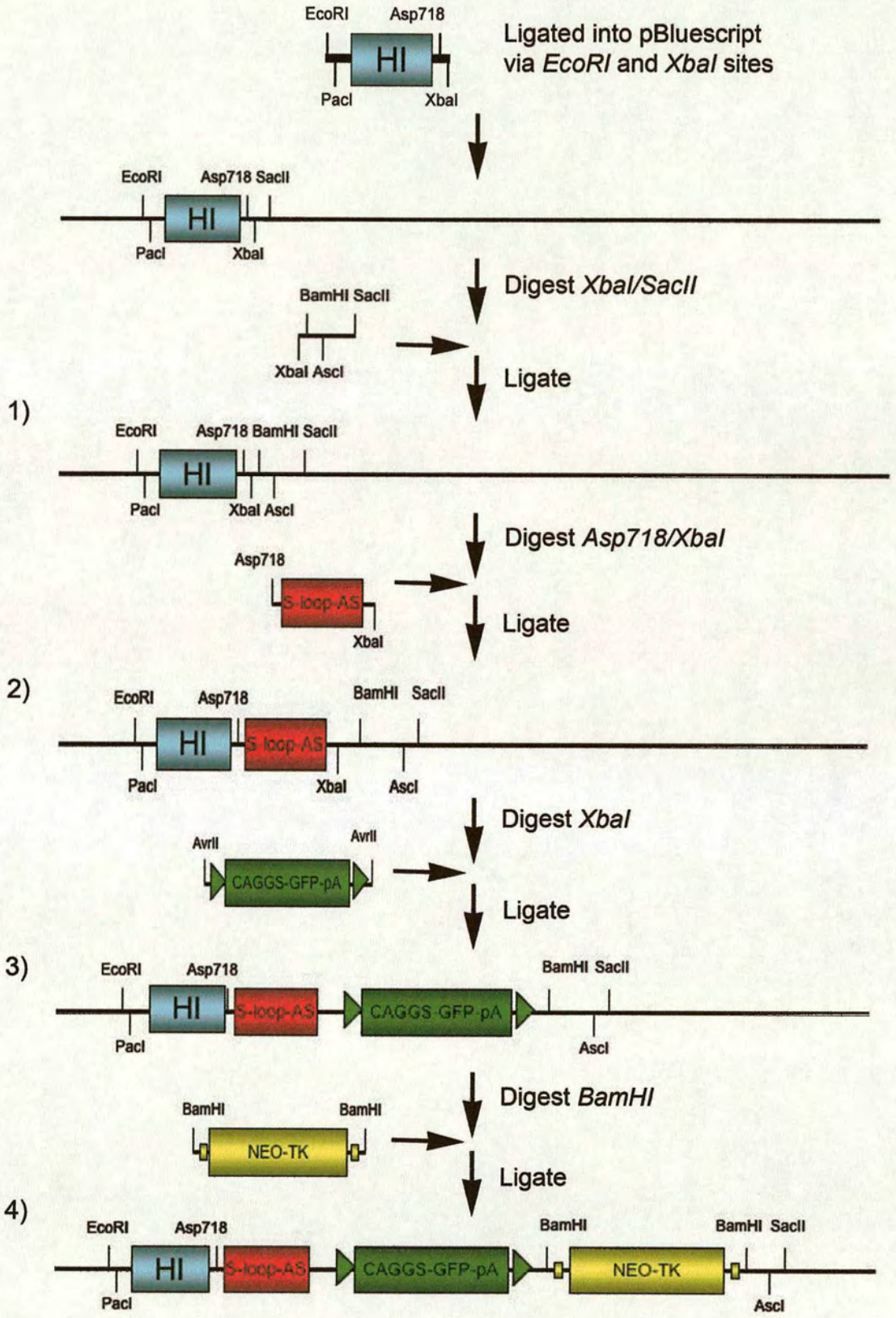


Figure 6.4 See over for legend.

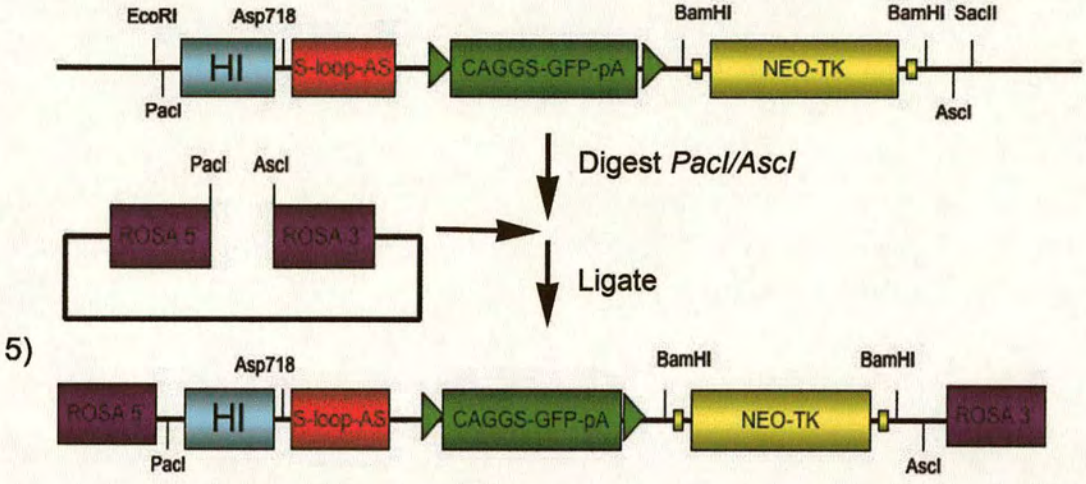


Figure 6.4 Strategy for construction a targeting cassette that will ubiquitously express shRNA from the ROSA26 locus. Construct contains: human H1 promoter (H1), sense-loop-antisense sequence encoding a *Rhox4* shRNA (S-loop-AS), floxed CAGGS-GFP cassette (CAGGFP), *frt*-site flanked neomycin-thymidine kinase selection cassette (NEO-TK) and 5' and 3' ROSA26 locus homology arms (ROSA 5' and ROSA 3'). Restriction sites are indicated.

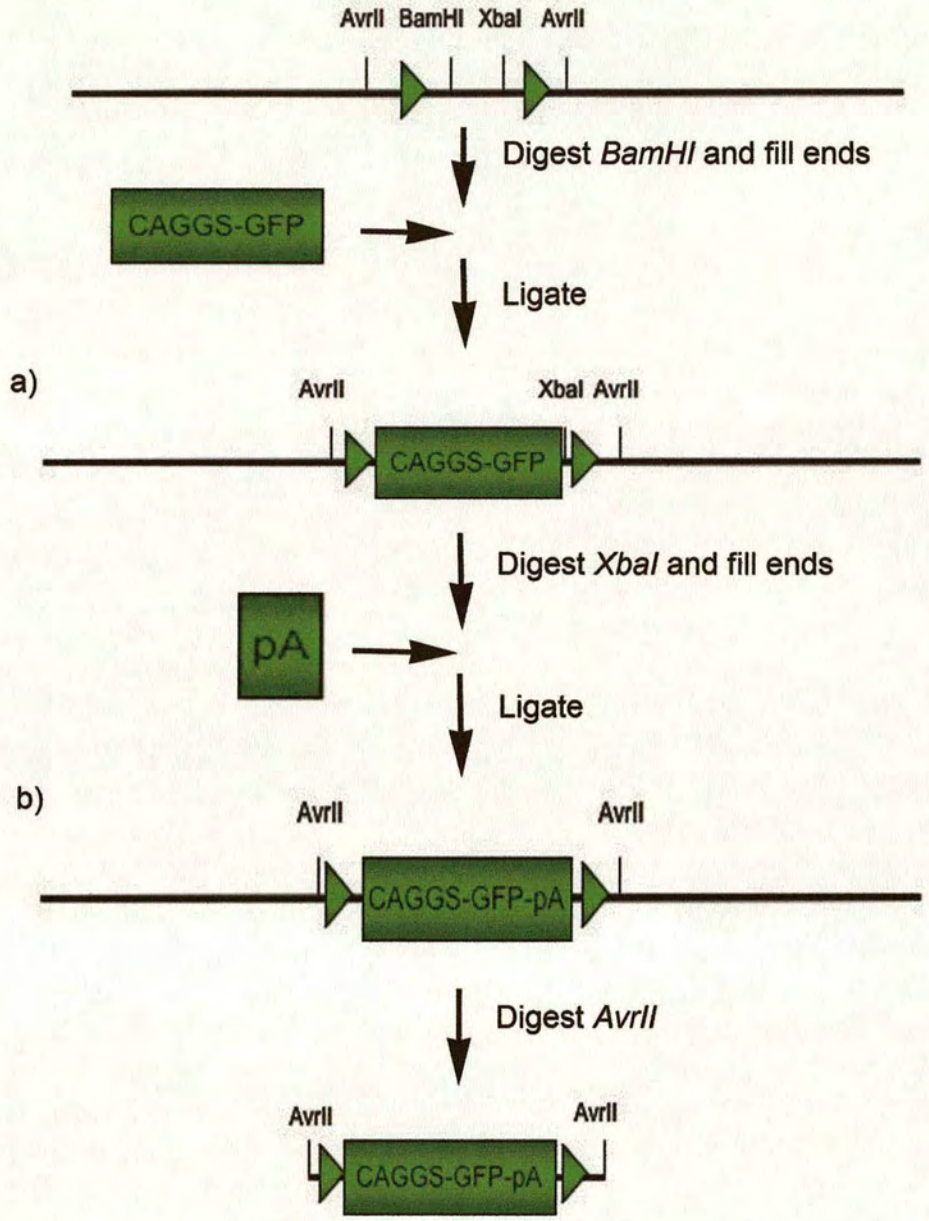


Figure 6.5 Cloning strategy used for constructing a loxP flanked GFP over-expression construct. Construct contains loxP sites (triangles), CAGGS-GFP and a polyadenylation site with a transcriptional pause (pA). Restriction sites are indicated.

6.2.3 Generation of a conditional shRNA expression construct for *Rhox4* knockdown

In addition to expression in the thymus, *Rhox4* expression is found in the placenta, yolk sac and testis. Furthermore, *Rhox4* been proposed to play a role in ES cell differentiation, as ES cell colonies expressing high levels of full-length *Rhox4* antisense RNA failed to differentiation in the absence of leukaemia inhibitory factor (Jackson, Baird et al. 2002; Jackson, Baird et al. 2003). These data suggest that a strategy for *Rhox4* knockdown based on ubiquitous expression of shRNA would result in preimplantation lethality. A recently described *Cre*-inducible strategy (Fritsch, Martinez et al. 2004) was therefore employed to make further targeting constructs expressing the *Rhox4* shRNAs. Here, insertion of a floxed CAGGS-GFP cassette between the sense and antisense components results in termination of transcription from the HI promoter (due to the presence of a run of five thymines in the CAGGS promoter). However, after *Cre*-mediated excision, the remaining *loxP* site acts as the loop sequence between the sense and antisense strand and transcription of the shRNA is initiated.

To introduce the LoxP-CAGGS-GFP-LoxP cassette between the sense and antisense strands of the shRNA, constructs were built as for the non-*Cre*-inducible constructs (Figure 6.6) with the alterations in the following steps:

2) The oligonucleotides containing the shRNAs contained an *AvrII* site in place of the loop sequence to allow insertion of *loxP-CAGGS-GFP-loxP*.
3) The *loxP-CAGGS-GFP-loxP* was inserted in between the sense and antisense strands of the DNA encoding the shRNA using the *AvrII* sites in both backbone and insert.

After determining which RNAi sequences mediate efficient *Rhox4* knockdown, both the inducible and non-inducible constructs will be introduced into the ROSA26 locus by standard homologous recombination and validated for expression of shRNA by Northern blotting. Demonstration of functional knockdown by the ES cell clones

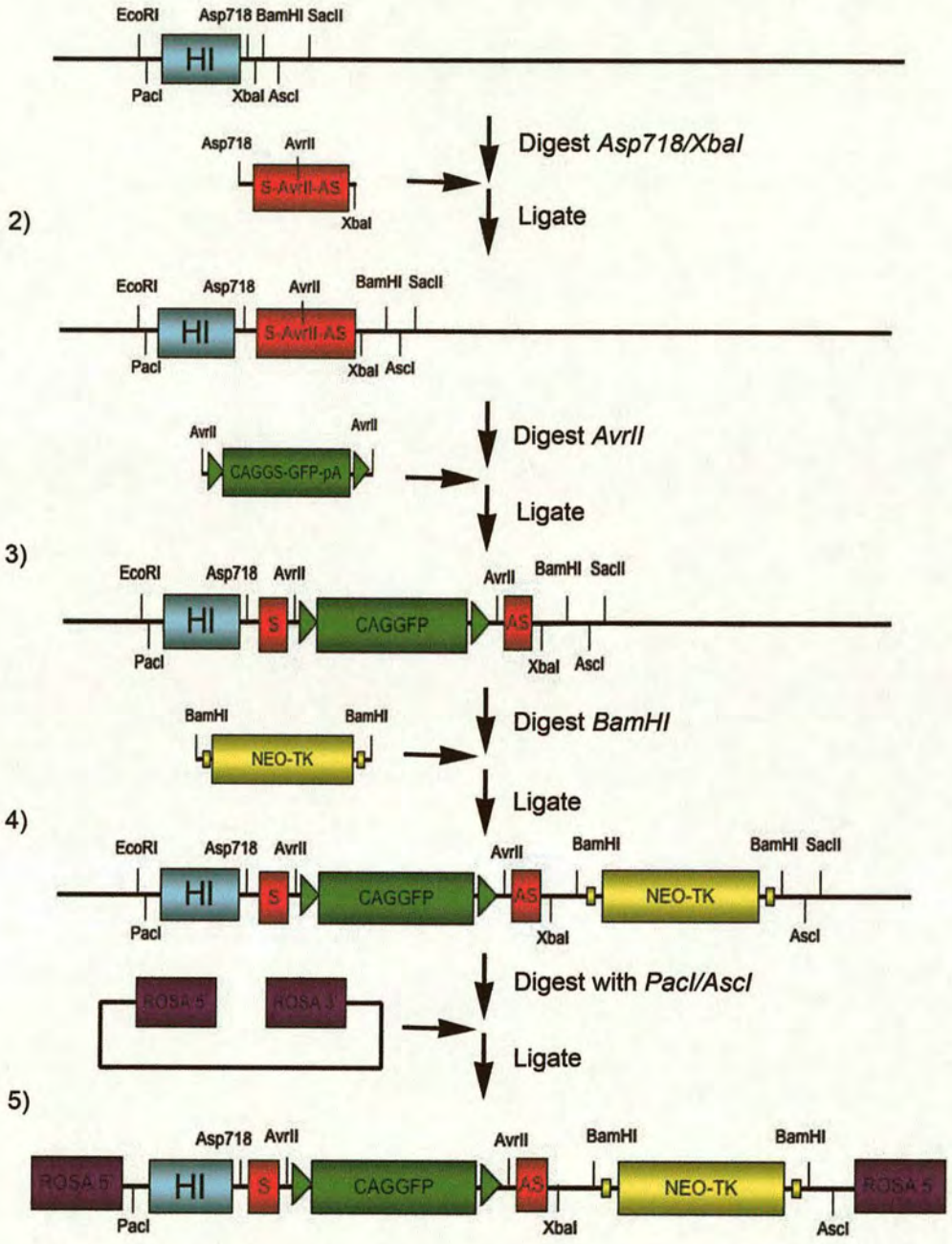


Figure 6.6 Strategy for construction of a targeting cassette that will ubiquitously express shRNA from the ROSA26 locus after *Cre*-mediated excision of GFP. Construct contains: human H1 promoter (H1); sense-*AvrII*-antisense sequence encoding a 19bp sense (S) and antisense (AS) *Rhox4* sequence separated by an *AvrII* site which will code for an *Rhox4* shRNA after *Cre*-mediated excision of GFP, floxed CAGGS-GFP cassette (CAGGFP); *frt*-site flanked neomycin-thymidine kinase selection cassette (NEO-TK); and 5' and 3' ROSA26 locus homology arms (ROSA 5' and ROSA 3'). Steps 1 and 2 illustrated in Figure 6.2. Restriction sites are indicated.

will require either the use of ES cell differentiation protocols that favour the production of cells of the endoderm lineage (Kubo, Shinozaki et al. 2004; Tada, Era et al. 2005), or transient transfection of clones with the *Rhox4* over-expression construct.

6.3 Discussion

In this chapter, I have described construction of both non-inducible and *Cre*-inducible constructs designed to mediate knockdown of *Rhox4* in mice, by expression of shRNA from the ROSA26 locus. Initial attempts to validate the RNAi sequences chosen for this approach were unsuccessful, due to the low level of *Rhox4* expression in available cell lines. I further show that the level of *Rhox4* expression after an ES cell differentiation protocol is not sufficient to allow this technique to be used for demonstrating functional knockdown of *Rhox4* in ES cells containing introduced shRNA expression constructs. Therefore, an alternative validation strategy, based on construction of a stable *Rhox4* over-expressing cell line has been initiated.

The RNAi pathway plays an essential role in mammalian development as shown by the early embryonic lethal phenotype of *Dicer*$^{-/-}$ mice (Bernstein, Kim et al. 2003), and also plays an important role in anti-viral immunity (Li, Li et al. 2002). The presence of dsRNA in a cell triggers activation of dsRNA-activated protein kinase (Williams 1999) and the production of type 1 interferons resulting in rapid suppression of translation and cell death (Kerr and Brown 1978; Lee and Esteban 1994). Although it was initially believed that these responses were only activated by dsRNA greater than 80bp (Bevilacqua and Cech 1996), and therefore that shRNA used for mediating RNAi did not activate these pathways, it is now clear that this is not the case (Bridge, Pebernard et al. 2003). None of the analyses of shRNA-expressing transgenic mice reported to date have looked for activation of anti-viral pathways or other off-target effects. Whilst several reports have shown that shRNA knockdown mice phenocopy the knockout phenotype, indicating specificity (Carmell, Zhang et al. 2003; Kunath, Gish et al. 2003; Coumoul, Shukla et al. 2005), they do not exclude the possibility that off-target effects were also present. However, mice expressing shRNA and mediating effective RNAi can be generated both from oocytes and ES cells, the mice have been reported to be fertile and transgene expression is maintained in offspring (Carmell, Zhang et al. 2003; Coumoul, Shukla et al. 2005).

Designing appropriate controls to account for off-target effects is not simple. The strategy described here includes a control in which the shRNA contains an antisense sequence that should be ineffective at mediating RNAi, which controls for the presence for dsRNA. Ideally, an additional control would be the demonstration that the induced phenotype could be rescued by *Rhox4* expression, to control for activation of the RNAi pathway. In this case, the *Rhox4* control would have to be mutated to protect it from RNAi-mediated destruction; in the specific case of *Rhox4,* construction of this control may also entail analysis of upstream region(s) of *Rhox4* copies required for correct expression. Several approaches to reducing off-target effects have been suggested, including use of RNA Polymerase II promoters - since the high expression levels driven by RNA Polymerase III promoters may exacerbate any non-specific effects, and RNA Polymerase II promoters produce lower levels of expression. Furthermore, unlike RNA Polymerase III promoters, examples of tissue specific promoters are known. With respect to non-specific effects, *in silico* analysis of the potential for sequence-specific off-target knockdown suggests that the use of a complementary 21-nucleotide sequence is optimal. In designing *Rhox4*-specific RNAis, a 19bp sequence was chosen as the HI promoter used requires a 21bp sequence including two uracils at the 3' end of the antisense strand; since no *Rhox4* sequence that was conserved between copies and predicted to be effective at mediating RNAi was found that contained two thymidines at the 3' end, a 21bp sequence could not be designed. The use of longer, 29bp shRNAs has also been suggested as these appear to be more potent at mediating RNAi (Kim, Behlke et al. 2005; Siolas, Lerner et al. 2005).

After successful introduction of the shRNA expression cassette to the ROSA26 locus and before generating lines of transgenic mice, it is necessary to demonstrate efficient expression of the shRNA and *Rhox4* knockdown. Although data reported here confirmed that ES cells do express *Rhox4,* as previously reported (Jackson, Baird et al. 2002), they do so at a barely detectible level, preventing the demonstration of successful knockdown in undifferentiated ES cells. *Rhox4* was

reported to be upregulated during *in vitro* ES cell differentiation using embryoid bodies (Jackson, Baird et al. 2002). Although the data presented here also demonstrate *Rhox4* upregulation using this differentiation protocol, the real level of *Rhox4* expression does not reach a level sufficiently above background to enable screening for efficient knockdown due to variability between replicates. *Rhox4* has been reported to play a role in ES cell differentiation to all lineages (Jackson, Baird et al. 2002), as demonstrated by the presence of small ES cell colonies that fail to differentiate in the absence of leukaemia inhibitory factor after over-expression of a full length *Rhox4* antisense transcript. Since *Rhox4* is expressed at a very low level at all stages of ES cell differentiation, these data suggest that ES cells are very sensitive to the level of *Rhox4* expression, with implications for generating *Rhox4*-specific RNAi ES cells. However, over-expression of the full-length (893bp) *Rhox4* anti-sense transcript described will have resulted in the presence of long pieces of dsRNA, which, as discussed above, will generate a strong anti-viral response. Since Jackson *et al* did not provide controls to show that the ES cell phenotype resulted from knockdown of *Rhox4*, it is possible that the reported phenotype was at least in part due to the presence of dsRNA. Construction of *Rhox4*-specific RNAi ES cells that mediate efficient knockdown may therefore help to address whether the ES cell phenotype described is due to knockdown of *Rhox4*. However, as discussed, these studies will need to demonstrate complete mRNA knockdown, control for the presence of shRNA, and rescue of phenotype by expression of a mutated *Rhox4*.

Concluding remarks

In this thesis, I have investigated the expression of *Rhox4* in thymus organogenesis and the structure of the *Rhox4* locus. These data show that *Rhox4* marks the ventral region of the 3rd pharyngeal pouch endoderm, which will form the thymic primordium, before the onset of *Foxn1* expression. This expression pattern suggested that *Rhox4* might play a role in specification of the thymus. *Hoxa3*, *Gcm2* and *Foxn1* have established roles in formation of the common thymus and parathyroid primordia. However, no change in *Rhox4* expression was found in mouse embryos null for these genes, indicating that none of these genes is a regulator of *Rhox4,* and that regulation of *Rhox4* expression in the thymus lies outside of established transcription factor pathways.

Rhox4 is a member of the *Rhox* family of homeobox-containing genes present in three clusters on the mouse X chromosome. Analysis of the *Rhox4* locus lead to identification of a previously unreported duplication containing a tandem array of three family members, *Rhox2, 3* and *4*. Eight complete copies of *Rhox2* and *3* and seven of *Rhox4* are present in this array. Multiple *Rhox4* copies were present in all sub-species of mice investigated, suggesting that the duplications arose before their divergence. Detailed analysis of *Rhox4* copy expression showed that all copies are expressed, although some differences between tissues were observed. Recent work has suggested that *Rhox* gene expression in the developing testis is under the control of co-linear mechanisms. Due to the expression of *Rhox4* at all stages of thymus development, this analysis was expanded to the expression of the *Rhox* α cluster genes during thymus organogenesis. No evidence was found to indicate that co-linear expression mechanisms were operating. In fact, of the mouse α cluster only *Rhox4* was expressed in the thymus. The predicted rat *Rhox4* was present in a single copy in the rat genome, suggesting that the duplications present in mice might have occurred at the time of divergence of mice and rats and, due to the reproductive expression of these genes, the duplication events may have played a role in the speciation event. Surprisingly *Rhox3,* but not *Rhox4* was expressed in all stages of rat thymus

development investigated, suggesting that rat *Rhox3* might be the functional orthologue of mouse *Rhox4*. Although these data indicate that there is a role for an *Rhox* family member in the murine embryonic thymus, no expression of either human *Rhox* gene was identified in this tissue. Finally, an RNAi based strategy was constructed for investigating the role of *Rhox4* in mouse thymus organogenesis.

The identification of seven very highly conserved copies of *Rhox4* and eight copies of *Rhox2* and *3* in a tandem array on the mouse X chromosome is unprecedented. Especially since, with the exception of the most closely related species, the rat, no orthologues of *Rhox2, 3* or *4* could be identified. However, in nearly every detail the duplication events identified are a perfect example of duplication and evolution that has been talked about and modelled by molecular evolutionary biologists for decades.

The *Rhox* family of homeobox-containing genes are expressed in reproductive and extra-embryonic tissues (Maclean, Chen et al. 2005), low sequence identity and evidence for positive selection is often found in reproductive genes as exemplified by previous work on mouse and rat *Rhox5* (Sutton and Wilkinson 1997) and *Rhox3* and *4* (this thesis). The mouse genome sequence shows that families of genes involved in reproduction have undergone lineage-specific amplification, such that large gene families have single or few human homologues (Waterston, Lindblad-Toh et al. 2002). Again, the *Rhox* family is a good example of this; there are only two human *Rhox* genes (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002), whereas there are 12 mouse (Maclean, Chen et al. 2005) and almost certainly 12 rat genes (Sutton and Wilkinson 1997), this thesis and M.F. Wilkinson, sequences submitted to NCBI) and additionally a further tandem duplication event has occurred in the mouse lineage resulting in seven *Rhox4* and eight *Rhox2* and *3* copies.

Although difficult to prove definitively, rapid evolution of reproductive and extra-embryonic genes coupled with lineage-specific duplications is very likely to play a

primary role in establishing and enforcing species barriers. The tandem duplications present within the *Rhox* α cluster are only found in the mouse locus, since evidence was only found for the existence of a single *Rhox4* copy in rat. As the duplications were present in all domestic and wild species of mice investigated, they could have occurred at the time of the divergence of these two lineages and therefore may have played a role in this event. Furthermore, the tandem array may have had a role in the divergence of sub-species of mice, assuming that there was a single duplication event that took place before their divergence. Although the C57BL/6 gene duplicates are highly identical, there is evidence that this is due to gene conversion, and the duplicates may well have diverged significantly from the ancestral duplicate gene. The degree of divergence of the different sub-species of mice can easily be addressed by sequence and expression analysis, and this would shed light upon the time of the duplication event or events and the trajectories of the duplicates since that time. For example it would be interesting to determine whether there is evidence for rapid divergence, which would suggest that the genes have played a role in the generation and maintenance of different sub-species.

It is unclear whether there is a conserved role for an *Rhox* gene in the thymus, since expression is found in both mouse (this thesis) (Jackson, Baird et al. 2002; Maclean, Chen et al. 2005) and rat (this thesis) but no thymic expression of human family members has been detected (this thesis) (Wayne, MacLean et al. 2002). This suggests that either thymic expression evolved after the division of human and murine lineages but before mouse and rats divided, or that expression in the thymus has been lost in humans: there is currently no evidence to indicate which hypothesis may be correct. Since gene duplication provides an opportunity for evolutionary novelty, including novel expression patterns, expression in the thymus may have evolved after the duplication event that resulted in the production of 12 murine *Rhox* family members. Determination of the role of *Rhox4* in the mouse thymus, establishing whether rat *Rhox3* expression during thymus organogenesis matches that of mouse *Rhox4* and gaining a better understanding of thymus development in human will aid in addressing this question. It is interesting to note that although mouse *Rhox7* is expressed in the stomach and *Rhox8* in the intestine (Maclean, Chen

et al. 2005), no expression of the human genes has been reported in these tissues either (Geserick, Weiss et al. 2002; Wayne, MacLean et al. 2002). Also, at least co-linear expression of *Rhox* family members is likely to have evolved after separation of the human and murine lineages, due to their likely divergence from a single common ancestor.

Rhox4 expression during mouse embryogenesis initially marks a broad region of the foregut endoderm at E8.5; by E9.5 expression is restricted to the ventral region of the 3rd pharyngeal pouch, from which the thymus arises. Expression is down-regulated after E10.5 and from E11.25, *Foxn1,* which is required for further thymus organogenesis, is expressed in its place. This striking restriction in expression suggested the hypothesis that *Rhox4* may play a role in specification of the thymic domain of the 3rd pharyngeal pouch, or in pharyngeal pouch development. Very little is know about the evolutionary origin of the thymus, however, thymi are found in zebrafish, where they appear similar to mammalian thymi with the exception of their ectopic location (Boehm, Bleul et al. 2003). Is it therefore likely that *Rhox4* really does play a role in specification of the mouse thymus? Treatment of *C. elegans* with RNAi libraries covering 86% of the genome found that in only 10.3% of knockdown genes resulted in a phenotype, and of these over half were lethal (Castillo-Davis and Hartl 2003). Genes that contained a single orthologue in another species were much more likely to give a phenotype, suggesting that genes with paralogues, and species-specific genes, are not essential for viability but are responsible for the differences between species. Mining of expression databases gave a similar result, orthologues that have recently undergone duplication events are unlikely to have strongly correlated expression patterns and conserved roles (Huminiecki and Wolfe 2004). The suggestion is therefore, that *Rhox4* is most likely to play a mouse specific role in the thymus. Morphological differences between human and mouse thymic primordia are present after primordia formation (A. Farley, unpublished) (Weller 1933; Norris 1938). These include shape, which is narrow and enlongated in human but spherical in mouse, and structure – a visible lobular structure is present in human but not in mouse. However, *Rhox4* is first expressed before primordia formation, suggesting that early differences in the primordium may also exist. Given the widespread use of

mouse as a model system for studying cellular processes and disease, understanding the differences between mouse and man may be as important as the similarities

References

Abu-Issa, R., G. Smyth, et al. (2002). "Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse." Development **129**(19): 4613-25.

Achaz, G., P. Netter, et al. (2001). "Study of intrachromosomal duplications among the eukaryote genomes." Mol Biol Evol **18**(12): 2280-8.

Adham, I. M., M. Gille, et al. (2005). "The scoliosis (sco) mouse: a new allele of Pax1." Cytogenet Genome Res **111**(1): 16-26.

Adkins, R. M., E. L. Gelke, et al. (2001). "Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes." Mol Biol Evol **18**(5): 777-91.

Aharoni, A., L. Gaidukov, et al. (2005). "The 'evolvability' of promiscuous protein functions." Nat Genet **37**(1): 73-6.

Ahn, B. Y., K. J. Dornfeld, et al. (1988). "Effect of limited homology on gene conversion in a Saccharomyces cerevisiae plasmid recombination system." Mol Cell Biol **8**(6): 2442-8.

Akam, M. (1989). "Hox and HOM: homologous gene clusters in insects and vertebrates." Cell **57**(3): 347-9.

Alexander, J., M. Rothenberg, et al. (1999). "casanova plays an early and essential role in endoderm formation in zebrafish." Dev Biol **215**(2): 343-57.

Anderson, G., K. L. Anderson, et al. (1997). "Fibroblast dependency during early thymocyte development maps to the CD25+ CD44+ stage and involves interactions with fibroblast matrix molecules." Eur J Immunol **27**(5): 1200-6.

Anderson, G. and E. J. Jenkinson (2001). "Lymphostromal interactions in thymic development and function." Nat Rev Immunol **1**(1): 31-40.

Anderson, G., E. J. Jenkinson, et al. (1993). "MHC class II-positive epithelium and mesenchyme cells are both required for T-cell development in the thymus." Nature **362**(6415): 70-3.

Anderson, M., S. K. Anderson, et al. (2000). "Thymic vasculature: organizer of the medullary epithelial compartment?" Int Immunol **12**(7): 1105-10.

Ashfield, R., A. J. Patel, et al. (1994). "MAZ-dependent termination between closely spaced human complement genes." Embo J **13**(23): 5656-67.

Bachiller, D., J. Klingensmith, et al. (2003). "The role of chordin/Bmp signals in mammalian pharyngeal development and DiGeorge syndrome." Development **130**(15): 3567-78.

Balciunaite, G., M. P. Keller, et al. (2002). "Wnt glycoproteins regulate the expression of FoxN1, the gene defective in nude mice." Nat Immunol **3**(11): 1102-8.

Baldini, A. (2005). "Dissecting contiguous gene defects: TBX1." Curr Opin Genet Dev **15**(3): 279-84.

Baxter, R. M. and J. L. Brissette (2002). "Role of the nude gene in epithelial terminal differentiation." J Invest Dermatol **118**(2): 303-9.

Bennett, A. R., A. Farley, et al. (2002). "Identification and characterization of thymic epithelial progenitor cells." Immunity **16**(6): 803-14.

Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-6.

Bernstein, E., S. Y. Kim, et al. (2003). "Dicer is essential for mouse development." Nat Genet **35**(3): 215-7.

Bevilacqua, P. C. and T. R. Cech (1996). "Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR." Biochemistry **35**(31): 9983-94.

Blackburn, C. C., C. L. Augustine, et al. (1996). "The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors." Proc Natl Acad Sci U S A **93**(12): 5742-6.

Blackburn, C. C. and N. R. Manley (2004). "Developing a new paradigm for thymus organogenesis." Nat Rev Immunol **4**(4): 278-89.

Bleul, C. C. and T. Boehm (2005). "BMP Signaling Is Required for Normal Thymus Development." J Immunol **175**(8): 5213-21.

Bockman, D. E. and M. L. Kirby (1984). "Dependence of thymus development on derivatives of the neural crest." Science **223**(4635): 498-500.

Boehm, T., C. C. Bleul, et al. (2003). "Genetic dissection of thymus development in mouse and zebrafish." Immunol Rev **195**: 15-27.

Bonhomme, F. and J.-L. Guenet (1996). "The laboratory mouse and its wild relatives." Lyon MF, Rastan S, Brown SDM (eds) Genetic variants and strains of the laboratory mouse, Oxford University Press: pp1577-1582.

Bonini, N. M., W. M. Leiserson, et al. (1993). "The eyes absent gene: genetic control of cell survival and differentiation in the developing Drosophila eye." Cell **72**(3): 379-95.

Bray, N., I. Dubchak, et al. (2003). "AVID: A global alignment program." Genome Res **13**(1): 97-102.

Bridge, A. J., S. Pebernard, et al. (2003). "Induction of an interferon response by RNAi vectors in mammalian cells." Nat Genet **34**(3): 263-4.

Carmell, M. A., L. Zhang, et al. (2003). "Germline transmission of RNAi in mice." Nat Struct Biol **10**(2): 91-2.

Castillo-Davis, C. I. and D. L. Hartl (2003). "Conservation, relocation and duplication in genome evolution." Trends Genet **19**(11): 593-7.

Chambon, P. (1996). "A decade of molecular biology of retinoic acid receptors." Faseb J **10**(9): 940-54.

Chapman, D. L., N. Garvey, et al. (1996). "Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development." Dev Dyn **206**(4): 379-90.

Cheyette, B. N., P. J. Green, et al. (1994). "The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system." Neuron **12**(5): 977-96.

Chiaromonte, F., W. Miller, et al. (2003). "Gene length and proximity to neighbors affect genome-wide expression levels." Genome Res **13**(12): 2602-8.

Chisaka, O. and M. R. Capecchi (1991). "Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5." Nature **350**(6318): 473-9.

Chisaka, O. and Y. Kameda (2005). "Hoxa3 regulates the proliferation and differentiation of the third pharyngeal arch mesenchyme in mice." Cell Tissue Res **320**(1): 77-89.

Chun, J. Y., Y. J. Han, et al. (1999). "Psx homeobox gene is X-linked and specifically expressed in trophoblast cells of mouse placenta." Dev Dyn **216**(3): 257-66.

Condie, B. G. and M. R. Capecchi (1993). "Mice homozygous for a targeted disruption of Hoxd-3 (Hox-4.1) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis." Development **119**(3): 579-95.

Cordier, A. C. and S. M. Haumont (1980). "Development of thymus, parathyroids, and ultimo-branchial bodies in NMRI and nude mice." Am J Anat **157**(3): 227-63.

Couly, G., S. Creuzet, et al. (2002). "Interactions between Hox-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head." Development **129**(4): 1061-73.

Couly, G., A. Grapin-Botton, et al. (1998). "Determination of the identity of the derivatives of the cephalic neural crest: incompatibility between Hox gene expression and lower jaw development." Development **125**(17): 3445-59.

Coumoul, X., V. Shukla, et al. (2005). "Conditional knockdown of Fgfr2 in mice using Cre-LoxP induced RNA interference." Nucleic Acids Res **33**(11): e102.

Crossley, P. H. and G. R. Martin (1995). "The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo." Development **121**(2): 439-51.

Crump, J. G., L. Maves, et al. (2004). "An essential role for Fgfs in endodermal pouch formation influences later craniofacial skeletal patterning." Development **131**(22): 5703-16.

Crump, J. G., M. E. Swartz, et al. (2004). "An integrin-dependent role of pouch endoderm in hyoid cartilage development." PLoS Biol **2**(9): E244.

Deschamps, J. and J. van Nes (2005). "Developmental regulation of the Hox genes during axial morphogenesis in the mouse." Development **132**(13): 2931-42.

Deutsch, U., G. R. Dressler, et al. (1988). "Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development." Cell **53**(4): 617-25.

Dietrich, S. and P. Gruss (1995). "undulated phenotypes suggest a role of Pax-1 for the development of vertebral and extravertebral structures." Dev Biol **167**(2): 529-48.

Dupe, V., N. B. Ghyselinck, et al. (1999). "Key roles of retinoic acid receptors alpha and beta in the patterning of the caudal hindbrain, pharyngeal arches and otocyst in the mouse." Development **126**(22): 5051-9.

Dykxhoorn, D. M. and J. Lieberman (2005). "The silent revolution: RNA interference as basic biology, research tool, and therapeutic." Annu Rev Med **56**: 401-23.

Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.

Elliott, B., C. Richardson, et al. (1998). "Gene conversion tracts from double-strand break repair in mammalian cells." Mol Cell Biol **18**(1): 93-101.

Emes, R. D., L. Goodstadt, et al. (2003). "Comparison of the genomes of human and mouse lays the foundation of genome zoology." Hum Mol Genet **12**(7): 701-9.

Ferretti, E., H. Marshall, et al. (2000). "Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins." Development **127**(1): 155-66.

Ferrier, D. E. and C. Minguillon (2003). "Evolution of the Hox/ParaHox gene clusters." Int J Dev Biol **47**(7-8): 605-11.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.

Flanagan, S. P. (1966). "'Nude', a new hairless gene with pleiotropic effects in the mouse." Genet Res **8**(3): 295-309.

Force, A., W. A. Cresko, et al. (2005). "The origin of subfunctions and modular gene regulation." Genetics **170**(1): 433-46.

Force, A., M. Lynch, et al. (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-45.

Frank, J., C. Pignata, et al. (1999). "Exposing the human nude phenotype." Nature **398**(6727): 473-4.

Fritsch, L., L. A. Martinez, et al. (2004). "Conditional gene knock-down by CRE-dependent short interfering RNAs." EMBO Rep **5**(2): 178-82.

Galtier, N. (2003). "Gene conversion drives GC content evolution in mammalian histones." Trends Genet **19**(2): 65-8.

Geserick, C., B. Weiss, et al. (2002). "OTEX, an androgen-regulated human member of the paired-like class of homeobox genes." Biochem J **366**(Pt 1): 367-75.

Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**(6982): 493-521.

Gill, J., M. Malin, et al. (2002). "Generation of a complete thymic microenvironment by MTS24(+) thymic epithelial cells." Nat Immunol **3**(7): 635-42.

Gordon, J., A. R. Bennett, et al. (2001). "Gcm2 and Foxn1 mark early parathyroid- and thymus-specific domains in the developing third pharyngeal pouch." Mech Dev **103**(1-2): 141-3.

Gordon, J., V. A. Wilson, et al. (2004). "Functional evidence for a single endodermal origin for the thymic epithelium." Nat Immunol **5**(5): 546-53.

Gotter, J. and B. Kyewski (2004). "Regulating self-tolerance by deregulating gene expression." Curr Opin Immunol **16**(6): 741-5.

Graham, A. and A. Smith (2001). "Patterning the pharyngeal arches." Bioessays **23**(1): 54-61.

Gruss, P. and C. Walther (1992). "Pax in development." Cell **69**(5): 719-22.

Gunther, T., Z. F. Chen, et al. (2000). "Genetic ablation of parathyroid glands reveals another source of parathyroid hormone." Nature **406**(6792): 199-203.

Guo, L., L. Degenstein, et al. (1996). "Keratinocyte growth factor is required for hair development but not for wound healing." Genes Dev **10**(2): 165-75.

Halder, G., P. Callaerts, et al. (1998). "Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development." Development **125**(12): 2181-91.

Halder, G., P. Callaerts, et al. (1995). "New perspectives on eye evolution." Curr Opin Genet Dev **5**(5): 602-9.

Hammond, S. M., E. Bernstein, et al. (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells." Nature **404**(6775): 293-6.

Hammond, W. S. (1954). "Origin of thymus in the chick embryo." J Morphol **95**: 501-613.

Han, Y. J., A. R. Park, et al. (1998). "Psx, a novel murine homeobox gene expressed in placenta." Gene **207**(2): 159-66.

Haworth, K. E., C. Healy, et al. (2004). "Regionalisation of early head ectoderm is regulated by endoderm and prepatterns the orofacial epithelium." Development **131**(19): 4797-806.

He, X. and J. Zhang (2005). "Gene complexity and gene duplicability." Curr Biol **15**(11): 1016-21.

Heanue, T. A., R. Reshef, et al. (1999). "Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation." Genes Dev **13**(24): 3231-43.

Hemberger, M. (2002). "The role of the X chromosome in mammalian extra embryonic development." Cytogenet Genome Res **99**(1-4): 210-7.

Hetzer-Egger, C., M. Schorpp, et al. (2002). "Thymopoiesis requires Pax9 function in thymic epithelial cells." Eur J Immunol **32**(4): 1175-81.

Hogan, B. L. (1999). "Morphogenesis." Cell **96**(2): 225-33.

Hogstrand, K. and J. Bohme (1997). "Gene conversion of major histocompatibility complex genes in the mouse spermatogenesis is a premeiotic event." Mol Biol Cell **8**(12): 2511-7.

Hollander, G. A., B. Wang, et al. (1995). "Developmental control point in induction of thymic cortex regulated by a subpopulation of prothymocytes." Nature **373**(6512): 350-3.

Hughes, M. J. and D. W. Andrews (1997). "A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system." FEBS Lett **414**(1): 19-22.

Huminiecki, L. and K. H. Wolfe (2004). "Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse." Genome Res **14**(10A): 1870-9.

Ivins, S., K. Lammerts van Beuren, et al. (2005). "Microarray analysis detects differentially expressed genes in the pharyngeal region of mice lacking Tbx1." Dev Biol **285**(2): 554-69.

Jackson, M., J. W. Baird, et al. (2002). "Cloning and characterisation of EHOX, a novel homeobox gene essential for ES cell differentiation." J Biol Chem.

Jackson, M., J. W. Baird, et al. (2003). "Expression of a novel homeobox gene Ehox in trophoblast stem cells and pharyngeal pouch endoderm." Dev Dyn **228**(4): 740-4.

Jacobs, Y., C. A. Schnabel, et al. (1999). "Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity." Mol Cell Biol **19**(7): 5134-42.

Jameson, S. C., K. A. Hogquist, et al. (1995). "Positive selection of thymocytes." Annu Rev Immunol **13**: 93-126.

Janes, S. M., T. A. Ofstad, et al. (2004). "Transient activation of FOXN1 in keratinocytes induces a transcriptional programme that promotes terminal differentiation: contrasting roles of FOXN1 and Akt." J Cell Sci **117**(Pt 18): 4157-68.

Jerome, L. A. and V. E. Papaioannou (2001). "DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1." Nat Genet **27**(3): 286-91.

Jiang, X., D. H. Rowitch, et al. (2000). "Fate of the mammalian cardiac neural crest." Development **127**(8): 1607-16.

Kameda, Y., Y. Arai, et al. (2004). "The role of Hoxa3 gene in parathyroid gland organogenesis of the mouse." J Histochem Cytochem **52**(5): 641-51.

Kameda, Y., T. Nishimaki, et al. (2002). "Homeobox gene hoxa3 is essential for the formation of the carotid body in the mouse embryos." Dev Biol **247**(1): 197-209.

Kang, Y. L., H. Li, et al. (2004). "A novel PEPP homeobox gene, TOX, is highly glutamic acid rich and specifically expressed in murine testis and ovary." Biol Reprod **70**(3): 828-36.

Kerr, I. M. and R. E. Brown (1978). "pppA2'p5'A2'p5'A: an inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells." Proc Natl Acad Sci U S A **75**(1): 256-60.

Khvorova, A., A. Reynolds, et al. (2003). "Functional siRNAs and miRNAs exhibit strand bias." Cell **115**(2): 209-16.

Kim, D. H., M. A. Behlke, et al. (2005). "Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy." Nat Biotechnol **23**(2): 222-6.

Klug, D. B., C. Carter, et al. (1998). "Interdependence of cortical thymic epithelial cell differentiation and T-lineage commitment." Proc Natl Acad Sci U S A **95**(20): 11822-7.

Klug, D. B., C. Carter, et al. (2002). "Cutting edge: thymocyte-independent and thymocyte-dependent phases of epithelial patterning in the fetal thymus." J Immunol **169**(6): 2842-5.

Kmita, M. and D. Duboule (2003). "Organizing axes in time and space; 25 years of colinear tinkering." Science **301**(5631): 331-3.

Kmita, M., N. Fraudeau, et al. (2002). "Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs." Nature **420**(6912): 145-50.

Kokubu, C., B. Wilm, et al. (2003). "Undulated short-tail deletion mutation in the mouse ablates Pax1 and leads to ectopic activation of neighboring Nkx2-2 in domains that normally express Pax1." Genetics **165**(1): 299-307.

Kopelman, N. M., D. Lancet, et al. (2005). "Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms." Nat Genet **37**(6): 588-9.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell **44**(2): 283-92.

Kozak, M. (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs." Nucleic Acids Res **15**(20): 8125-48.

Krumlauf, R. (1992). "Evolution of the vertebrate Hox homeobox genes." Bioessays **14**(4): 245-52.

Kubo, A., K. Shinozaki, et al. (2004). "Development of definitive endoderm from embryonic stem cells in culture." Development **131**(7): 1651-62.

Kunath, T., G. Gish, et al. (2003). "Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype." Nat Biotechnol **21**(5): 559-61.

Laclef, C., E. Souil, et al. (2003). "Thymus, kidney and craniofacial abnormalities in Six 1 deficient mice." Mech Dev **120**(6): 669-79.

Laky, K., L. Lefrancois, et al. (2000). "Enterocyte expression of interleukin 7 induces development of gammadelta T cells and Peyer's patches." J Exp Med **191**(9): 1569-80.

Le Douarin, N. M. and F. V. Jotereau (1975). "Tracing of cells of the avian thymus through embryonic life in interspecific chimeras." J Exp Med **142**(1): 17-40.

Lee, S. B. and M. Esteban (1994). "The interferon-induced double-stranded RNA-activated protein kinase induces apoptosis." Virology **199**(2): 491-6.

Lewis, E. B. (1978). "A gene complex controlling segmentation in Drosophila." Nature **276**(5688): 565-70.

Li, H., W. X. Li, et al. (2002). "Induction and suppression of RNA silencing by an animal virus." Science **296**(5571): 1319-21.

Lin, T. P., P. A. Labosky, et al. (1994). "The Pem homeobox gene is X-linked and exclusively expressed in extraembryonic tissues during early murine development." Dev Biol **166**(1): 170-9.

Lind, E. F., S. E. Prockop, et al. (2001). "Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development." J Exp Med **194**(2): 127-34.

Lindsay, E. A., F. Vitelli, et al. (2001). "Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice." Nature **410**(6824): 97-101.

Lynch, M. and J. S. Conery (2000). "The evolutionary fate and consequences of duplicate genes." Science **290**(5494): 1151-5.

Lynch, M. and A. Force (2000). "The probability of duplicate gene preservation by subfunctionalization." Genetics **154**(1): 459-73.

Lynch, M. and V. Katju (2004). "The altered evolutionary trajectories of gene duplicates." Trends Genet **20**(11): 544-9.

Macatee, T. L., B. P. Hammond, et al. (2003). "Ablation of specific expression domains reveals discrete functions of ectoderm- and endoderm-derived FGF8 during cardiovascular and pharyngeal development." Development **130**(25): 6361-74.

Maclean, J. A., 2nd, M. A. Chen, et al. (2005). "Rhox: a new homeobox gene cluster." Cell **120**(3): 369-82.

Makova, K. D. and W. H. Li (2003). "Divergence in the spatial pattern of gene expression between human duplicate genes." Genome Res **13**(7): 1638-45.

Mallon, A. M., L. Wilming, et al. (2004). "Organization and evolution of a gene-rich region of the mouse genome: a 12.7-Mb region deleted in the Del(13)Svea36H mouse." Genome Res **14**(10A): 1888-901.

Manley, N. R. and C. C. Blackburn (2003). "A developmental look at thymus organogenesis: where do the non-hematopoietic cells in the thymus come from?" Curr Opin Immunol **15**(2): 225-32.

Manley, N. R. and M. R. Capecchi (1995). "The role of Hoxa-3 in mouse thymus and thyroid development." Development **121**(7): 1989-2003.

Manley, N. R. and M. R. Capecchi (1997). "Hox group 3 paralogous genes act synergistically in the formation of somitic and neural crest-derived structures." Dev Biol **192**(2): 274-88.

Manley, N. R. and M. R. Capecchi (1998). "Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands." Dev Biol **195**(1): 1-15.

Manley, N. R., L. Selleri, et al. (2004). "Abnormalities of caudal pharyngeal pouch development in Pbx1 knockout mice mimic loss of Hox3 paralogs." Dev Biol **276**(2): 301-12.

Mansouri, A., G. Goudreau, et al. (1999). "Pax genes and their role in organogenesis." Cancer Res **59**(7 Suppl): 1707s-1709s; discussion 1709s-1710s.

Marais, G. (2003). "Biased gene conversion: implications for genome and sex evolution." Trends Genet **19**(6): 330-8.

Mignone, F., C. Gissi, et al. (2002). "Untranslated regions of mRNAs." Genome Biol **3**(3): REVIEWS0004.

Moore, M. A. and J. J. Owen (1967). "Experimental studies on the development of the thymus." J Exp Med **126**(4): 715-26.

Moore-Scott, B. A. and N. R. Manley (2005). "Differential expression of Sonic hedgehog along the anterior-posterior axis regulates patterning of pharyngeal

pouch endoderm and pharyngeal endoderm-derived organs." Dev Biol **278**(2): 323-35.

Muller, T. S., C. Ebensperger, et al. (1996). "Expression of avian Pax1 and Pax9 is intrinsically regulated in the pharyngeal endoderm, but depends on environmental influences in the paraxial mesoderm." Dev Biol **178**(2): 403-17.

Murti, J. R., M. Bumbulis, et al. (1992). "High-frequency germ line gene conversion in transgenic mice." Mol Cell Biol **12**(6): 2545-52.

Nehls, M., B. Kyewski, et al. (1996). "Two genetically separable steps in the differentiation of thymic epithelium." Science **272**(5263): 886-9.

Nehls, M., D. Pfeifer, et al. (1994). "New member of the winged-helix protein family disrupted in mouse and rat nude mutations." Nature **372**(6501): 103-7.

Neubuser, A., H. Koseki, et al. (1995). "Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1." Dev Biol **170**(2): 701-16.

Noden, D. M. (1983). "The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues." Dev Biol **96**(1): 144-65.

Noonan, J. P., J. Grimwood, et al. (2004). "Gene conversion and the evolution of protocadherin gene cluster diversity." Genome Res **14**(3): 354-66.

Norris, E. (1938). "The morphogenesis and histogenesis of the thymus gland in man; in which the origin of the Hassell's corpuscles of the human thymus is discovered." Contrib. Embryol. **27**: 191-221.

Nossal, G. J. (1994). "Negative selection of lymphocytes." Cell **76**(2): 229-39.

Ohno, S. (1970). "Evolution by gene and genome duplication." Springer-Verlag, New York.

Ohuchi, H., Y. Hori, et al. (2000). "FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development." Biochem Biophys Res Commun **277**(3): 643-9.

Orr-Weaver, T. L. and J. W. Szostak (1985). "Fungal recombination." Microbiol Rev **49**(1): 33-58.

Otting, G., Y. Q. Qian, et al. (1990). "Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution." Embo J **9**(10): 3085-92.

Peters, H., A. Neubuser, et al. (1998). "Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities." Genes Dev **12**(17): 2735-47.

Petes, T. D. and C. W. Hill (1988). "Recombination between repeated genes in microorganisms." Annu Rev Genet **22**: 147-68.

Prowse, D. M., D. Lee, et al. (1999). "Ectopic expression of the nude gene induces hyperproliferation and defects in differentiation: implications for the self-renewal of cutaneous epithelia." Dev Biol **212**(1): 54-67.

Quinlan, R., P. Martin, et al. (2004). "The role of actin cables in directing the morphogenesis of the pharyngeal pouches." Development **131**(3): 593-9.

Radding, C. M. (1978). "Genetic recombination: strand transfer and mismatch repair." Annu Rev Biochem **47**: 847-80.

Reifers, F., E. C. Walsh, et al. (2000). "Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar)." Development **127**(2): 225-35.

Revest, J. M., R. K. Suniara, et al. (2001). "Development of the thymus requires signaling through the fibroblast growth factor receptor R2-IIIb." J Immunol **167**(4): 1954-61.

Rossel, M. and M. R. Capecchi (1999). "Mice mutant for both Hoxa1 and Hoxb1 show extensive remodeling of the hindbrain and defects in craniofacial development." Development **126**(22): 5027-40.

Rozas, J. and R. Rozas (1999). "DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis." Bioinformatics **15**(2): 174-5.

Sambrook, J., E. Fritsch, et al. (1989). "Molecular Cloning - A Laboratory Manual."

Sasai, Y. and E. M. De Robertis (1997). "Ectodermal patterning in vertebrate embryos." Dev Biol **182**(1): 5-20.

Scambler, P. (2003). "Velocardofacial/DiGeorge syndrome." In: Cooper, D. (Ed), Nature Encyclopedia of the Human Genome. Macmillan, London.

Schreier, J. E. and J. L. Hamilton (1952). "An experimental study of the origin of the parathyroid and thymus glands in the chich." J. Exptl. Zool. **119**: 165-187.

Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.

Selleri, L., M. J. Depew, et al. (2001). "Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation." Development **128**(18): 3543-57.

Sharp, S. B., T. A. Kost, et al. (1989). "Regulation of chicken alpha and beta actin genes and their hybrids inserted into myogenic mouse cells." Gene **80**(2): 293-304.

Sharpe, P. and I. Mason (2005). "Molecular embryology: methods and protocols." Humana Press.

Shinohara, T. and T. Honjo (1997). "Studies in vitro on the mechanism of the epithelial/mesenchymal interaction in the early fetal thymus." Eur J Immunol **27**(2): 522-9.

Siolas, D., C. Lerner, et al. (2005). "Synthetic shRNAs as potent RNAi triggers." Nat Biotechnol **23**(2): 227-31.

Soriano, P. (1999). "Generalized lacZ expression with the ROSA26 Cre reporter strain." Nat Genet **21**(1): 70-1.

Springer, M. S., W. J. Murphy, et al. (2003). "Placental mammal diversification and the Cretaceous-Tertiary boundary." Proc Natl Acad Sci U S A **100**(3): 1056-61.

Su, D., S. Ellis, et al. (2001). "Hoxa3 and pax1 regulate epithelial cell death and proliferation during thymus and parathyroid organogenesis." Dev Biol **236**(2): 316-29.

Su, D. M. and N. R. Manley (2000). "Hoxa3 and pax1 transcription factors regulate the ability of fetal thymic epithelial cells to promote thymocyte development." J Immunol **164**(11): 5753-60.

Su, D. M. and N. R. Manley (2002). "Stage-specific changes in fetal thymocyte proliferation during the CD4-8- to CD4+8+ transition in wild type, Rag1-/-, and Hoxa3,Pax1 mutant mice." BMC Immunol **3**(1): 12.

Su, D. M., S. Navarre, et al. (2003). "A domain of Foxn1 required for crosstalk-dependent thymic epithelial cell differentiation." Nat Immunol **4**(11): 1128-35.

Suniara, R. K., E. J. Jenkinson, et al. (2000). "An essential role for thymic mesenchyme in early T cell development." J Exp Med **191**(6): 1051-6.

Sutton, K. A. and M. F. Wilkinson (1997). "Rapid evolution of a homeodomain: evidence for positive selection." J Mol Evol **45**(6): 579-88.

Tada, S., T. Era, et al. (2005). "Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture." Development **132**(19): 4363-74.

Taddei, I., M. Morishima, et al. (2001). "Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/del22q11 syndromes." Proc Natl Acad Sci U S A **98**(20): 11428-31.

Takasaki, N., R. McIsaac, et al. (2000). "Gpbox (Psx2), a homeobox gene preferentially expressed in female germ cells at the onset of sexual dimorphism in mice." Dev Biol **223**(1): 181-93.

Takasaki, N., T. Rankin, et al. (2001). "Normal gonadal development in mice lacking GPBOX, a homeobox protein expressed in germ cells at the onset of sexual dimorphism." Mol Cell Biol **21**(23): 8197-202.

Taylor, J. S., Y. Van de Peer, et al. (2001). "Genome duplication, divergent resolution and speciation." Trends Genet **17**(6): 299-301.

Thornton, K. and M. Long (2002). "Rapid divergence of gene duplicates on the Drosophila melanogaster X chromosome." Mol Biol Evol **19**(6): 918-25.

Ting, C. T., S. C. Tsaur, et al. (1998). "A rapidly evolving homeobox at the site of a hybrid sterility gene." Science **282**(5393): 1501-4.

Trokovic, N., R. Trokovic, et al. (2003). "Fgfr1 regulates patterning of the pharyngeal region." Genes Dev **17**(1): 141-53.

Tucker, A. S. and A. Lumsden (2004). "Neural crest cells provide species-specific patterning information in the developing branchial skeleton." Evol Dev **6**(1): 32-40.

Vallender, E. J. and B. T. Lahn (2004). "How mammalian sex chromosomes acquired their peculiar gene content." Bioessays **26**(2): 159-69.

Vallender, E. J., N. M. Pearson, et al. (2005). "The X chromosome: not just her brother's keeper." Nat Genet **37**(4): 343-5.

van Ewijk, W., B. Wang, et al. (1999). "Thymic microenvironments, 3-D versus 2-D?" Semin Immunol **11**(1): 57-64.

Veitch, E., J. Begbie, et al. (1999). "Pharyngeal arch patterning in the absence of neural crest." Curr Biol **9**(24): 1481-4.

Vermot, J., K. Niederreither, et al. (2003). "Decreased embryonic retinoic acid synthesis results in a DiGeorge syndrome phenotype in newborn mice." Proc Natl Acad Sci U S A **100**(4): 1763-8.

Vitelli, F., M. Morishima, et al. (2002). "Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways." Hum Mol Genet **11**(8): 915-22.

Vitelli, F., I. Taddei, et al. (2002). "A genetic link between Tbx1 and fibroblast growth factor signaling." Development **129**(19): 4605-11.

Wagner, G. P., C. Amemiya, et al. (2003). "Hox cluster duplications and the opportunity for evolutionary novelties." Proc Natl Acad Sci U S A **100**(25): 14603-6.

Wallin, J., H. Eibel, et al. (1996). "Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation." Development **122**(1): 23-30.

Wallin, J., Y. Mizutani, et al. (1993). "A new Pax gene, Pax-9, maps to mouse chromosome 12." Mamm Genome **4**(7): 354-8.

Wang, B., G. A. Hollander, et al. (1996). "Natural killer cell development is blocked in the context of aberrant T lymphocyte ontogeny." Int Immunol **8**(6): 939-49.

Wang, X. and J. Zhang (2004). "Rapid evolution of mammalian X-linked testis-expressed homeobox genes." Genetics **167**(2): 879-88.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Wayne, C. M., J. A. MacLean, et al. (2002). "Two novel human X-linked homeobox genes, hPEPP1 and hPEPP2, selectively expressed in the testis." Gene **301**(1-2): 1-11.

Weller, G. (1933). "Development of the thyroid, parathryoid and thymus glands in man." Contributions to embryology **141**: 95-139.

Wendling, O., C. Dennefeld, et al. (2000). "Retinoid signaling is essential for patterning the endoderm of the third and fourth pharyngeal arches." Development **127**(8): 1553-62.

Williams, B. R. (1999). "PKR; a sentinel kinase for cellular stress." Oncogene **18**(45): 6112-20.

Wolfe, K. H. and P. M. Sharp (1993). "Mammalian gene evolution: nucleotide sequence divergence between mouse and rat." J Mol Evol **37**(4): 441-56.

Wyckoff, G. J., W. Wang, et al. (2000). "Rapid evolution of male reproductive genes in the descent of man." Nature **403**(6767): 304-9.

Xu, H., F. Cerrato, et al. (2005). "Timed mutation and cell-fate mapping reveal reiterated roles of Tbx1 during embryogenesis, and a crucial function during segmentation of the pharyngeal system via regulation of endoderm expansion." Development **132**(19): 4387-95.

Xu, P. X., I. Woo, et al. (1997). "Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode." Development **124**(1): 219-31.

Xu, P. X., W. Zheng, et al. (2002). "Eya1 is required for the morphogenesis of mammalian thymus, parathyroid and thyroid." Development **129**(13): 3033-44.

Yamazaki, H., E. Sakata, et al. (2005). "Presence and distribution of neural crest-derived cells in the murine developing thymus and their potential for differentiation." Int Immunol **17**(5): 549-58.

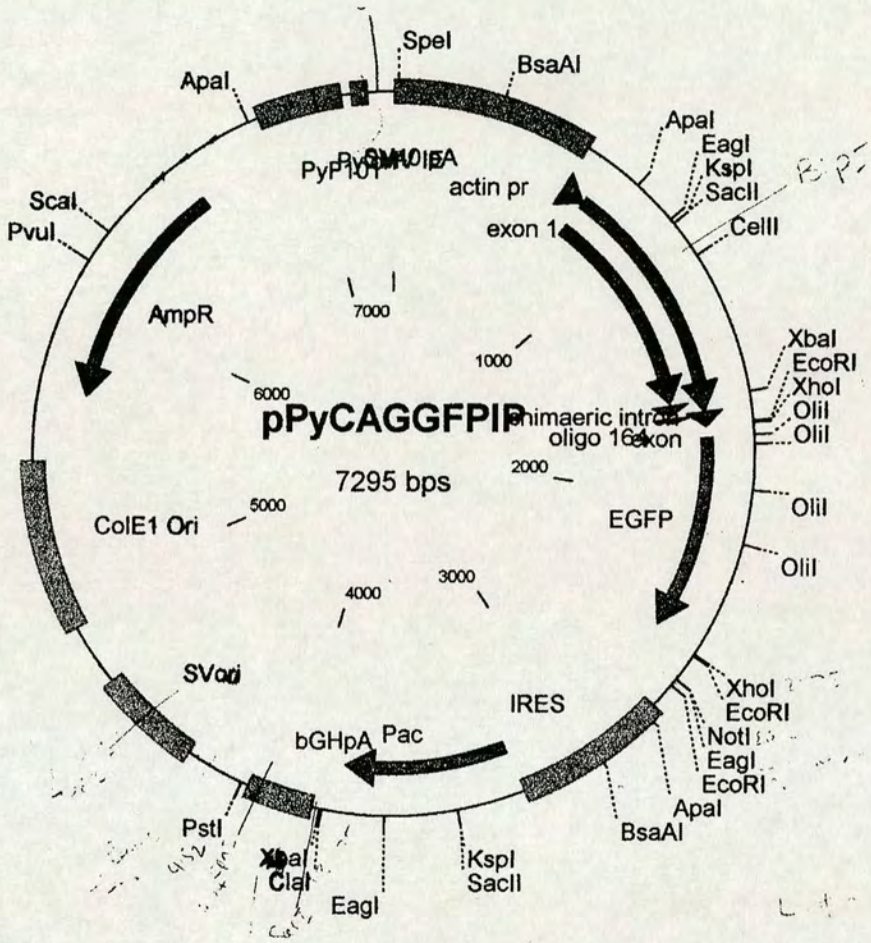
Zambrowicz, B. P., A. Imamoto, et al. (1997). "Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells." Proc Natl Acad Sci U S A **94**(8): 3789-94.

Zamisch, M., B. Moore-Scott, et al. (2005). "Ontogeny and regulation of IL-7-expressing thymic epithelial cells." J Immunol **174**(1): 60-7.

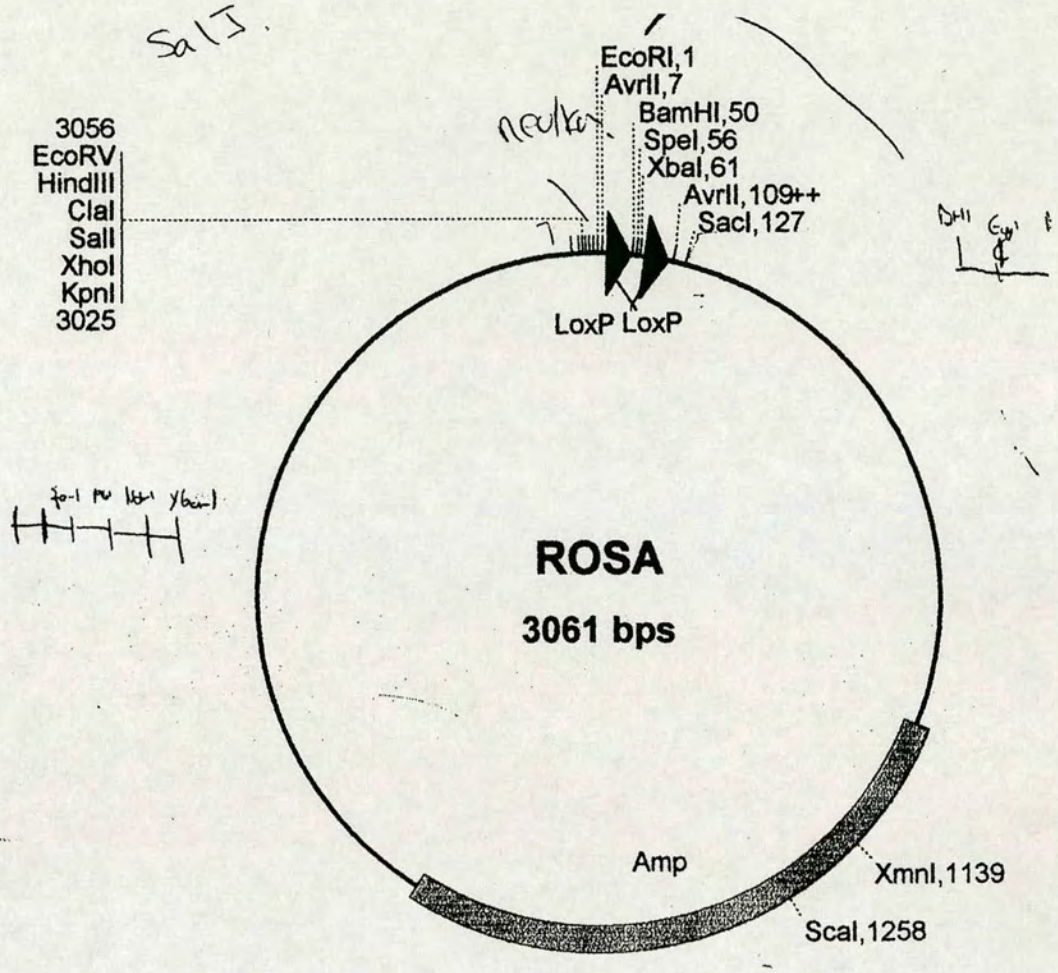
Zamore, P. D., T. Tuschl, et al. (2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." Cell **101**(1): 25-33.

Appendix

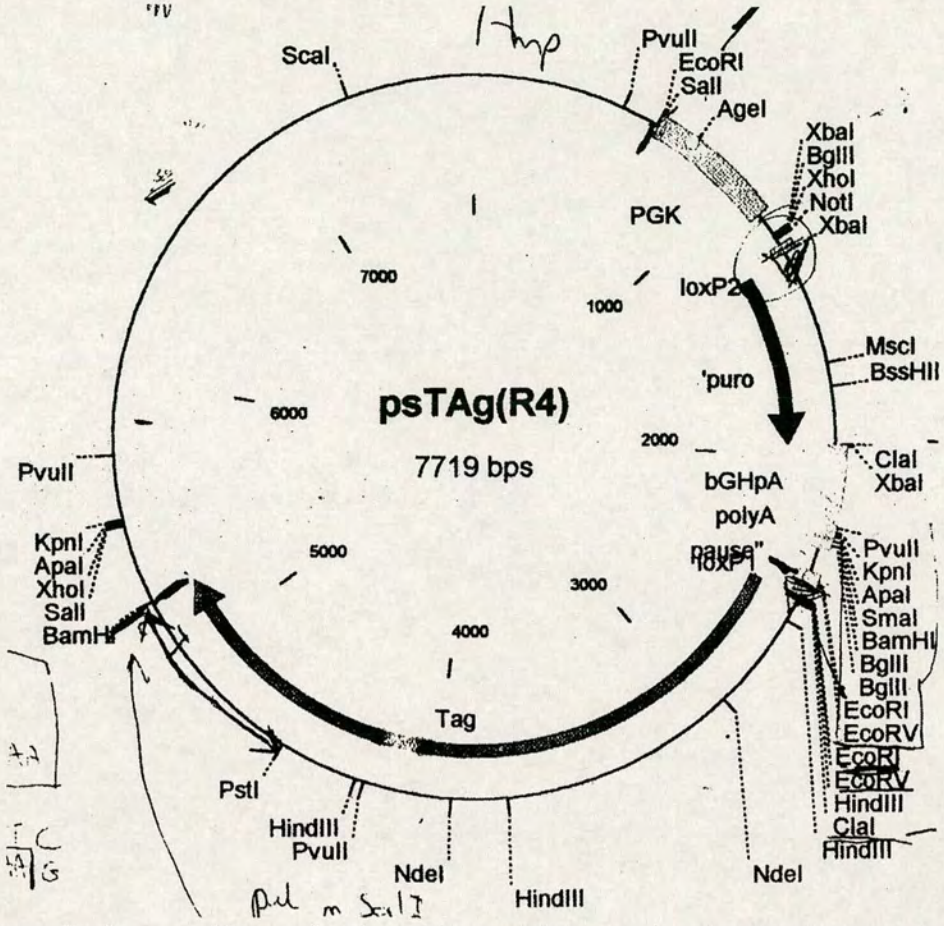
Maps of plasmids used in this thesis, no map was provided with the plasmid containing *frt-neo-tk-frt*:



Appendix 1: Map of plasmid used for a) construction of *Rhox4* over-expression construct. Full-length *Rhox4* was inserted in place of EGFP via digestion with *EcoRI*. The orientation of integration was determined by sequencing. b) EGFP was extracted from above plasmid via digestion with *EcoRI* and integrated into the shRNA-expressing targeting constructs as shown in Chapter 6.



Appendix 2 Plasmid containing two LoxP sites in the same orientation used for construction of a *floxed* EGFP. The EGFP was introduced into the *BamHI* site by blunt end cloning. A polyadenylation signal and transcriptional stop were introduced into the *XbaI* site by blunt end cloning.



Appendix 3 Map of plasmid containing a polyadenylation signal followed by a transcriptional stop. The polyadenylation and stop signals were removed by digestion with *XbaI/ EcoRI* and introduced into plasmid shown in Appendix 2.